

2005

Automated dental identification: A micro-macro decision-making approach

Diaa Eldin M. Nassar
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Nassar, Diaa Eldin M., "Automated dental identification: A micro-macro decision-making approach" (2005). *Graduate Theses, Dissertations, and Problem Reports*. 2348.
<https://researchrepository.wvu.edu/etd/2348>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Automated Dental Identification: A Micro-Macro Decision-Making Approach

Diaa Eldin M. Nassar

Dissertation submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Computer Engineering

Hany H. Ammar, Ph.D., Chair
Donald A. Adjero, Ph.D., Co-Chair
Xin Li, Ph.D.
Arun A. Ross, Ph.D.
Robert Howell, DDS.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2005

Keywords: Dental Identification, ADIS, Teeth Classification, Dental Chart, Hierarchical Information Fusion, Performance Limitations, and Copulas.

Copyright 2005 Diaa Eldin M. Nassar

ABSTRACT

Automated Dental Identification: A Micro-Macro Decision-Making Approach

Diaa Eldin M. Nassar

Identification of deceased individuals based on dental characteristics is receiving increased attention, especially with the large volume of victims encountered in mass disasters. In this work we consider three important problems in automated dental identification beyond the basic approach of tooth-to-tooth matching.

The first problem is on automatic classification of teeth into incisors, canines, premolars and molars as part of creating a data structure that guides tooth-to-tooth matching, thus avoiding illogical comparisons that inefficiently consume the limited computational resources and may also mislead the decision-making. We tackle this problem using principal component analysis and string matching techniques. We reconstruct the segmented teeth using the eigenvectors of the image subspaces of the four teeth classes, and then call the teeth classes that achieve least energy-discrepancy between the novel teeth and their approximations. We exploit teeth neighborhood rules in validating teeth-classes and hence assign each tooth a number corresponding to its location in a dental chart. Our approach achieves 82% teeth labeling accuracy based on a large test dataset of bitewing films.

Because dental radiographic films capture projections of distinct teeth; and often multiple views for each of the distinct teeth, in the second problem we look for a scheme that exploits teeth multiplicity to achieve more reliable match decisions when we compare the dental records of a subject and a candidate match. Hence, we propose a hierarchical fusion scheme that utilizes both aspects of teeth multiplicity for improving teeth-level (micro) and case-level (macro) decision-making. We achieve a genuine accept rate in excess of 85%.

In the third problem we study the performance limits of dental identification due to features capabilities. We consider two types of features used in dental identification,

namely teeth contours and appearance features. We propose a methodology for determining the number of degrees of freedom possessed by a feature set, as a figure of merit, based on modeling joint distributions using copulas under less stringent assumptions on the dependence between feature dimensions. We also offer workable approximations of this approach.

Dedication

To my family with love and appreciation,

To my mentors, and

To all those who tirelessly work for the best of all mankind.

Acknowledgements

I wish to express my gratitude to my advisor Professor Dr. Hany Ammar for his continuous support for me over many years of graduate studies; I very much appreciate his invaluable trust and best of all his great patience and tolerance. It has been my pleasure to work with a farsighted advisor like Professor Ammar.

I was privileged to have Dr. Donald Adjeroth co-advise my Ph.D. research, it has been a wonderful experience to get exposed to the great research qualities Dr. Adjeroth has, I highly appreciate his patience, humbleness and friendliness. Dr. Adjeroth's enthusiasm has always inspired me.

I am also very grateful to Dr. Xin Li and Dr. Arun Ross; not only for them being valuable members of my examination committee, but also for being skillful teachers. My gratitude also goes to Dr. Robert Howell for serving in my examination committee and for his help in developing a better understanding of the dentistry side of the problem.

It has been my honor to work with and study under great scholars, from whom I highly benefited. I am extremely grateful to Dr. Ali Mili, Dr. Ronald Klein, Dr. Anil Jain, Dr. Mohamed Abdel-Mottaleb, Dr. Matthew Valenti, and Dr. Katerina Goseva-Popstojanova.

I ought to thank all my WVU research colleagues: Dr. Gamal Fahmy, Satya Chekuri, Ayman Abaza, Eyad Haj-Said, Mythili Ogirila, Faisal Chaudhry, Mohamed Ikkery, Ali Bahu, Tayyeb Anwar, Usman Qureshi, Tianjian Wang, Lin Zhang, and Zainab Milwala. It has also been my pleasure to cooperate with Hong Chen from MSU, Omaira Nomair, Jindan Zhou, and Hossein Mahoor from University of Miami.

I am very much indebted to my great friends: Dr. Sherif Yacoub, Hicham Arafa, Tarek Abou-Zahra, Dr. Yasser Rasheed, Ahmed Badr, Mohab Kassem, Islam Attia, Waleed Farouk, Dr. Bassem Hamid, Dr. Iyad Muslih, Dr. Mohammad Al-Robee, Rabih Kreidly, Kalaivani Appukkutty, Rajesh Gunnalan, Walid Abdel-Moez, and Umasankar Kandaswamy for their great support and valuable advice.

My family has always been a source of endless support and love; I really cannot thank each of my family members enough. I am very thankful to my Father Mr. Mohammad A. Nassar, my Mother Mrs. Nadera El-Tawahini, my brother Belal and my sisters Aaya, Alaa, Afnan, and Asmaa. I am also very grateful to my Father-in-law Dr. Sherif M. El-Haggan and my Mother-in-law Mrs. Sohair Shoukri. I highly appreciate the generous attention given to my wife, my son and myself by Dr. Mostafa El-Haggan, and Dr. Ahmed M. El-Haggan and his family. And last but by no means least; I faithfully appreciate my great wife Doaa for her constant support, endurance, and high sense of responsibility; and my lovely son Omar for being a great source of joy and hope.

This research was supported under the US National Science Foundation (NSF) award number EIA-0131079 and the US National Institute of Justice (NIJ) award number 2001-RC-CX-K013.

Table of Contents

Chapter 1: Introduction	1
1.1 Postmortem Identification.....	1
1.2 The Automated Dental Identification System (ADIS).....	2
1.3 Dental Image Comparison	6
1.4 Research Objectives and Scope	7
1.5 Research Contributions.....	9
1.6 Dissertation Organization	9
Chapter 2: Background and The Problem.....	10
2.1 Background	10
2.2 Literature Review	12
2.2.1 Postmortem Dental Identification	12
2.2.2 Image Registration	19
2.2.3 Principal Component Analysis	22
2.2.3.1 EigenFaces, FisherFaces, and LaplacianFaces	24
2.2.4 Classifier Fusion	26
2.2.5 Performance Limits in Biometrics	32
2.3 Problem Statement.....	36
2.3.1 Problem One: Automatic Construction of Dental Charts	37
2.3.2 Problem Two: Micro and Macro Decision-Making.....	38
2.3.3 Problem Three: Performance Limits of Dental Identification	39
Chapter 3: Automatic Construction of Dental Charts.....	40
3.1 Overview	40
3.2 Teeth Reconstruction and Classification	41
3.2.1 The Eigenteeth	43
3.2.2 View Normalization.....	46

3.2.3 Projection and Classification	50
3.3 Class Validation and Number Assignment.....	51
3.4 Experimental Results.....	52
3.5 Discussion.....	59
3.6 Summary.....	61
Chapter 4: Micro and Macro Decision Making.....	63
4.1 Overview	63
4.2 Micro-Decision Making.....	65
4.3 Macro-Decision Making and Ranking.....	66
4.4 Experimental Results.....	70
4.5 Summary.....	75
Chapter 5: On Performance Limits of Dental Identification	77
5.1 Overview	77
5.2 Performance Limits of Dental Identification using Teeth Contours.....	79
5.2.1 Parameter Estimation	82
5.3 Performance of Dental Identification using appearance-based features.....	84
5.3.1 Estimating Limits on Performance under IID feature dimensions	85
5.3.2 Estimating the Number of Degrees of Freedom: A Generalized Approach	88
5.3.3 Estimating the Number of Degrees of Freedom: Approximations	90
5.4 Alternative Representations.....	93
5.5 Discussion.....	95
5.6 Summary.....	96
Chapter 6: Conclusions and Future Work.....	98
6.1 Research Conclusions	98
6.2 Future Work.....	100

Appendix: Tooth-to-Tooth Matching	103
A1 The Feature Extraction Layer.....	104
A2 The Bayesian Classification Layer.....	106
A3 Decision Making	107
A4 The Training Algorithm	108
A4.1 Training Objective	108
A4.2 Back Propagation.....	109
A4.3 Adaptive Uni-Variation of Weights.....	109
A4.4 A Hybrid Training Approach	111
A5 Experimental Results	112
A5.1 Non-learnable Parameters	113
A5.2 Training	113
A5.2.a Convergence	114
A5.2.b Examples of Training Experiments	115
A5.2.c An Interpretation of Feature Extraction Filters.....	116
A5.3 Testing.....	116
A5.3.a Testing with a Genuine Dataset	117
A5.3.b Testing with Synthetic Datasets.....	117
Bibliography	121

List of Figures

Figure 1: A block diagram of the prototype ADIS.	5
Figure 2: A block diagram of the image comparison component.....	7
Figure 3: Example of using dental charts for identification in WinID®.	16
Figure 4: Structure of a 32-cell, 4-quadrant dental chart.	37
Figure 5: Block diagram of the two stages of teeth classification.	41
Figure 6: Teeth reconstruction and initial classification.....	43
Figure 7: Sample of teeth used in constructing the image subspaces of the four teeth classes.	44
Figure 8: (a) Upward pointing crown (b) Conjugate of (a) under horizontal flipping.	46
Figure 9: Image representation of (a) <i>EigenIncisors</i> , (b) <i>EigenCanines</i> , (c) <i>EigenPremolars</i> , and (d) <i>EigenMolars</i>	46
Figure 10: The gradient vectors of (a) the test pattern, (b) a segmented tooth with upward crown, and (c) a segmented tooth with downward crown.....	48
Figure 11: Illustration of the orientation adjustment and ROI confinement steps of a segmented tooth.....	49
Figure 12: Image representation of the confined tooth ROI after resizing.	50
Figure 13: Reconstructed image of the view normalized molar (Figure 12) in the four subspaces.	51
Figure 14: Scaled error images between the molar (Figure 12) and its images in the four subspaces.	51
Figure 15: Teeth distribution of the test dataset used in teeth classification experiments.....	54
Figure 16: Examples of bitewing films (a – c) with correct teeth numbers (d – f) with incorrect teeth numbers.	57
Figure 17: Examples of lower periapical films (a, b) with correct teeth numbers (c, d) with incorrect teeth numbers.	57

Figure 18: Examples of upper periapical films (a, b) with correct teeth numbers (c, d) with incorrect teeth numbers.	58
Figure 19: Example of a rejected bitewing film due to inverted upper and lower quadrants.	58
Figure 20: Examples of initially misclassified teeth.....	60
Figure 21: A dental record with multiple films showing multiple views of teeth.	64
Figure 22: A hierarchical fusion scheme for micro and macro decision-making.....	65
Figure 23: Profile of the function $g_I(N_1, N_2, N_3) = \frac{(\sqrt{2} N_1 + N_3)^2}{64(N_1 + N_2 + N_3)}$	68
Figure 24: Profile of the function $g_I(0, N_2, N_3) = \frac{N_3^2}{64(N_2 + N_3)}$	68
Figure 25: Profile of the function $g_I(N_1, 0, N_3) = \frac{(\sqrt{2} N_1 + N_3)^2}{64(N_1 + N_3)}$	69
Figure 26: Profile of the function $g_I(N_1, N_2, 0) = \frac{N_1^2}{32(N_1 + N_2)}$	69
Figure 27: Distribution of the teeth-pair views of the test dataset.....	71
Figure 28: Distribution of case-to-case matching time.....	72
Figure 29: Hit rate based on fixed normalization of the ranking score.	74
Figure 30: Hit rate based on variable normalization of the ranking score.....	74
Figure 31: Examples of subject records (a and c) and their respective matches (b and d), ranked first in the match lists of these subjects, and a subject record (e) with a false accept record (f) ranked top of the match list of that subject.	75
Figure 32: A rectangular box bounding a tooth, also tolerance areas around its contour points	80
Figure 33: Example of manual matching of teeth-pairs using manually extracted contours..	83
Figure 34: Cumulative probability distribution of point matching distance using 641 points.	84

Figure 35: Average Hamming distance distribution for (a) Premolars, and (b) Molars.....	87
Figure 36: Binomial fitting of the average Hamming distance distributions for (a) Premolars, and (b) Molars.	88
Figure 37: The pyramidal architecture for image matching.	104
Figure 38: Block diagram of the feature extraction and classification stages.	104
Figure 39: The adaptive uni-variation of weights in pseudo code.....	111
Figure 40: The hybrid-training algorithm in pseudo code.	112
Figure 41: Some training pairs (a) unmatched pair, (b) matched pair.	114
Figure 42: Convergence profile of the matcher using the hybrid algorithm in section A4.4	115
Figure 43: Example of a ROI (a) and its extracted feature images; $Z_{ROI}^{[1]}$ (b) and $Z_{ROI}^{[2]}$ (c).	116
Figure 44: Performance of the tooth-to-tooth matcher under scaling variations.	119
Figure 45: Performance of the tooth-to-tooth matcher under rotation variations.	120

List of Tables

Table 1: The number of films and teeth used in testing of teeth classification.	54
Table 2: Confusion matrix for teeth classification in all films	55
Table 3: Confusion matrix for teeth classification in bitewing films only	56
Table 4: Confusion matrix for teeth classification in lower periapical films only	56
Table 5: Confusion matrix for teeth classification in upper periapical films only	56
Table 6: Error rates for genuine matches at the macro decision level.	72
Table 7: Sample statistics for the average Hamming distance of premolars and molars.	87
Table 8: Variation of performance limits with decision thresholds.....	88
Table 9: Examples of training experiments for the tooth-to-tooth matcher.....	116
Table 10: Results of testing the matcher using a genuine dataset.....	117
Table 11: Results of testing the matcher using synthetic datasets.....	119

Chapter 1: Introduction

Forensic Odontology is a branch of forensic science that studies identification of human individuals based on their dental features. The importance of dental features as postmortem biometric identifiers stem from their survivability and resistance to early decay. Automating the process of dental identification became inevitable due to the increased number of victims in mass disasters. In this chapter we overview the problem of automating the dental identification process, we identify the key components of an Automated Dental Identification System (ADIS), and we define the scope of this dissertation.

1.1 Postmortem Identification

Law enforcement agencies have exploited biometric identifiers, for decades, as a means for forensic identification. With the huge volume of cases that need to be investigated by forensic specialists and the evolution in information technology, it has become important to automate forensic identification systems.

Post-Mortem (PM) identification, i.e. identification after death, is a more difficult problem than Ante-Mortem (AM) identification, since few biometrics can be used. Postmortem identification is carried out using either positive or presumptive identification methods. Positive identification methods involve comparison of antemortem and postmortem data that are unique to an individual, while presumptive methods include identification based on “*visual recognition, personal effects, serology, anthropometric data, and medical history*” [1]. Positive PM identification methods include: “(i) *dental comparisons*, (ii) *comparisons of fingerprints, palm prints, or footprints*, (iii) *DNA identification*, and (iv) *radiographic superimposition*” [1]. Presumptive identification predominantly provides means for exclusion of potential mismatches based on race, gender, age, and blood type [1].

Under severe circumstances, such as those encountered in high energy mass disasters (e.g. airplane crashers, Tsunami, hurricanes, wars, etc.) or if identification is being attempted more than a couple of weeks after death (e.g. late discovery of corpses), most physiological biometrics (e.g. fingerprints) do not qualify as basis for identification. Under such circumstances, soft tissues of the human body would have decayed to unidentifiable status.

Therefore, a PM biometric identifier must outlive the early decay that affects body tissues [1][2].

Because of their survivability, diversity and availability, the best candidates for biometric PM identification are dental features. Forensic odontology is the branch of forensics that studies identification of human individuals based on their dental features. Forensic odontology utilizes three major areas: “(i) *diagnostic and therapeutic examination of injuries of jaws, teeth, and soft oral tissues*, (ii) *identification of individuals in criminal investigations and mass disasters*, and (iii) *identification and examination of bite marks*.” [1].

In PM identification, forensic odontologists rely mainly on dental radiographs, among other types of records e.g. oral photographs, denture models, and CAT scans, to compare the morphology of dental restorations (e.g. fillings, and crowns) of the unidentified persons to those of candidates in the missing persons file. With the significant improvement in the dental hygiene of the contemporary generations, and the deployment of some “*relatively radiolucent*” materials in fillings and restorations [3]; it is becoming important to shift to identification decisions based on inherent dental features like root and crown morphologies, teeth sizes, rotations, inter-teeth spacing and sinus patterns [2][4].

Manual radiograph comparison is a highly time-consuming process that requires high levels of skill and accuracy. With the increased volumes of both dental records and victims, the task of the forensic odontologists - which is matching PM to AM records - becomes tedious, more difficult and time consuming. Hence, computer-aided dental record comparison systems become the proper means for manipulating large volumes of data while maintaining accuracy, consistency, and low running cost [1][5].

1.2 The Automated Dental Identification System (ADIS)

In 1997, the Criminal Justice Information Services Division (CJIS) of the FBI created a dental task force (DTF) whose goal is to improve the utilization and effectiveness of the National Crime Information Center's (NCIC) Missing and Unidentified Persons (MUP) files. The DTF recommended the creation of a Digital Image Repository (DIR) and an Automated Dental Identification System (ADIS) with goals and objectives similar to the Automated Fingerprint Identification System (AFIS) but using dental characteristics instead of

fingerprints. The proclaimed ADIS is a highly automated system that when fed with raw subject dental records, will find a minimum set of candidate (or reference) records, ideally one, that have high similarities to the subject. Then, a forensic expert will examine the radiographs of the few candidates to make a final decision on the identity of the missing or unidentified person.

The Digital Image Repository (DIR) will contain dental images of patients and is linked to NCIC's MUP files, which contain non-image information such as age, gender, race, and blood type. This information is used to exclude candidates with impossible matches, thus reducing the search space.

In [6] we proposed a prelude research prototype of ADIS, which we later extended and refined in collaboration with CJIS experts - as depicted in Figure 1 - to become the conceptual framework of the ADIS prototype. ADIS is composed of two main components: A system for search and retrieval based on potential similarities and a verification system for matching based on low level comparison of dental images [7]. Given a subject record, identification is carried out as follows:

- The subject record is preprocessed in order to facilitate the subsequent processing steps. Record preprocessing involves the tasks of: (a) record cropping into dental films [8], (b) enhancement of films to compensate for possible poor contrast [9], (c) classification of films into bitewing, periapical, or panoramic views [10], (d) segmentation of teeth from films [10][11][12][13], and (e) annotating teeth with labels corresponding their location [14][15].
- High-level features are extracted from the preprocessed record and are used by the potential matches search component to query the DIR for reference records that possess high similarity to the subject, in high-level features sense. The potential matches search component then generates a candidate list of these records. Candidates are the bearers of reference records with dental/non-dental features that are potentially similar to those possessed by the bearer of the subject record. Initially, only dental high-level features (e.g. number/type of teeth), which are automatically extracted from dental records, are exploited in this stage. The potential matches search may optionally consider the outcome of searching the National Crime

Information Center (NCIC) repository using non-dental features, which include age, race, and gender.

- Finally, dental radiographs of the subject record are compared to those of each candidate. The outcome of the comparison is a short match list ranked according to the probability of match between the subject record and each qualifying candidate record.

The philosophy behind architecting the prototype ADIS as such is that the potential matches search component is a fast, high-recall retrieval system, while the image comparison component is a high-precision matching system. We envision the works presented by *Jain* and *Chen* in [11][12][16], *Nomair* and *Abdel-Mottaleb* in [13], and by *Zhou* and *Abdel-Mottaleb* in [10] as candidate realizations of the potential matches search component. These are complementary to our work, which mainly addresses the problem of instantiating the second component, namely the image comparison component.

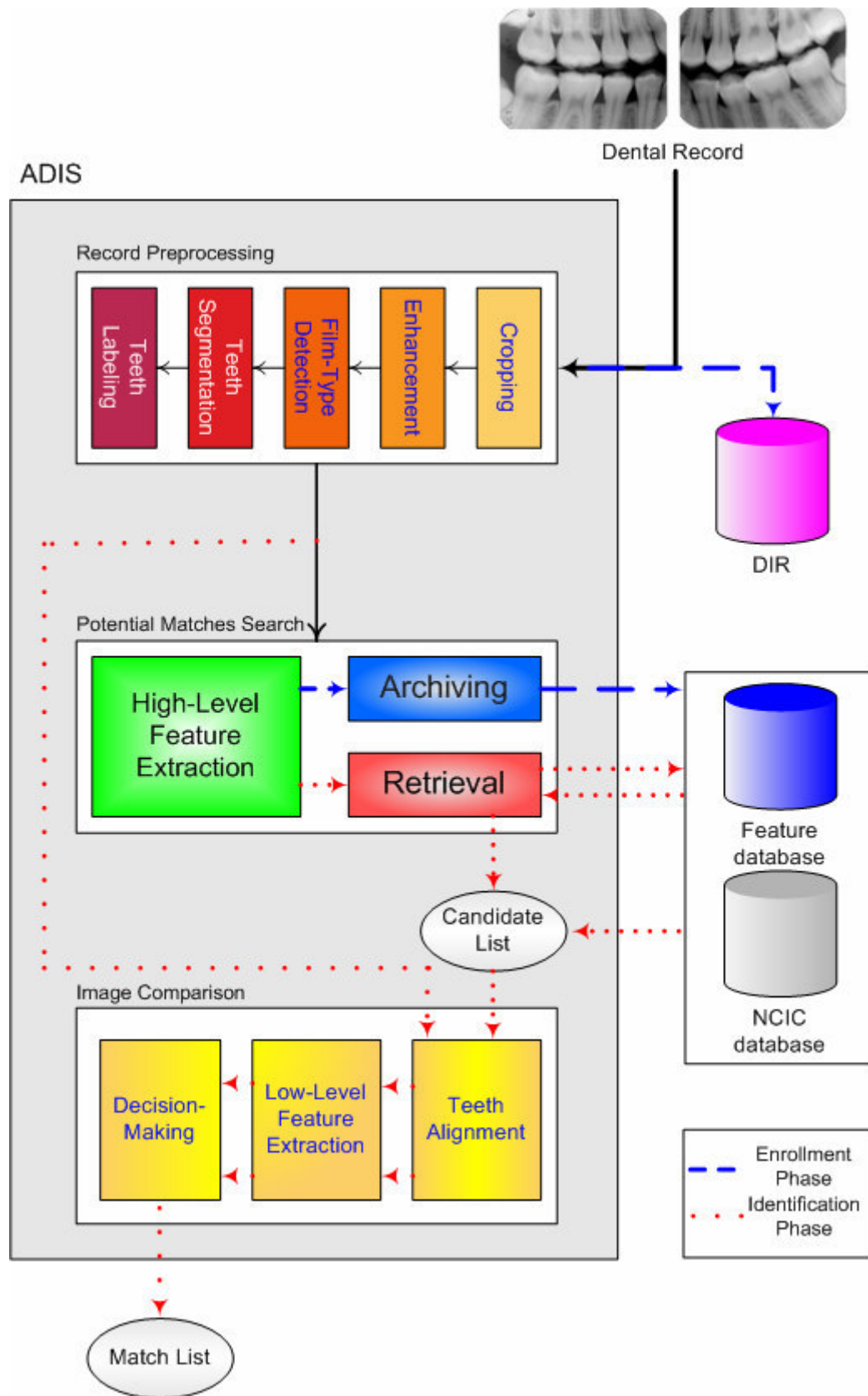


Figure 1: A block diagram of the prototype ADIS.

1.3 Dental Image Comparison

In Figure 2, we show a block diagram of the image comparison component. Guided by the output of the teeth-labeling step of the record-preprocessing component, we construct dental charts for subject and candidate records. A dental chart is a data structure that associates each segmented tooth with a cell in a dental atlas corresponding to the 32 possible teeth of an adult.

For a subject/candidate record-pair, we conduct pair-wise region of interest (ROI) comparisons, i.e. we compare a subject ROI to a corresponding reference ROI. The outcome of each ROI-pair comparison is a number representing the probability of a match between the particular ROI-pair. We select the teeth-pairs to compare based on the dental charts of the reference and subject records to avoid illogical comparisons (e.g. molars are compared to molars but not to canines).

Given an ROI-pair, we start with a hypothesis that the two objects are matched, at least based on teeth location. And accordingly we apply appropriate transformations that restore major geometric discrepancies between the ROI-pair, this is achieved by teeth alignment [17][18][19][20].

As a dental record usually comprises multiple films that may show more than a single view of a given tooth (e.g. the upper right first molar), we exploit this view multiplicity of a tooth in reaching a more robust decision about the match status of a subject/reference tooth-pair. We refer to the process of determining the match status of a subject/reference tooth-pair based on comparison of multiple views of this tooth-pair as micro-decision making.

Finally, with many micro-decisions (up to 32 in the case of a fully developed adult), we combine these micro-decisions into a macro-decision that determines the match status of the subject/candidate record-pair and accordingly whether the candidate record should be placed on the match list. We also provide a ranking score to sort the match list.

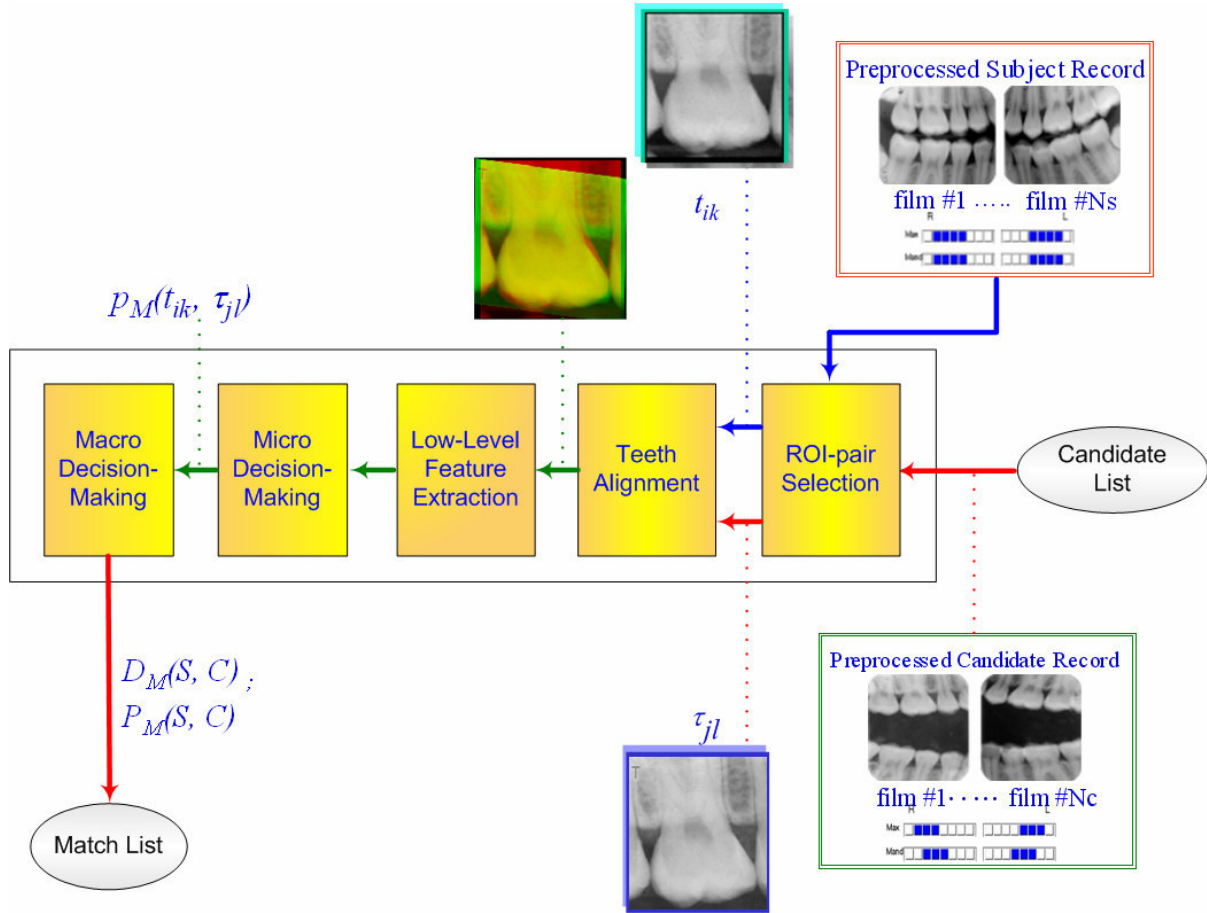


Figure 2: A block diagram of the image comparison component

1.4 Research Objectives and Scope

This research is broadly oriented towards developing computationally feasible high performance techniques to facilitate automatic identification of missing, wanted and unidentified persons using their dental characteristics recorded on digitized radiographic films. The research problems we study are part of the integral research framework of ADIS supported under the US National Science Foundation (NSF) award number EIA-0131079 and the US National Institute of Justice (NIJ) award number 2001-RC-CX-K013.

The research prototype of ADIS is required to timely achieve high identification accuracy; therefore each component of ADIS inherits these requirements. In this dissertation we will limit our scope to the following problems:

- a. **Teeth classification:** in selecting teeth pairs to compare, we want to avoid illogical pairing of reference and subject ROIs, as they are primarily illogical to compare.

Such comparisons inefficiently consume limited computational resources. Clearly match probabilities due to inappropriate comparisons could negatively affect the accuracy of macro decisions. Dentists categorize teeth into four classes: incisors, canines, premolars, and molars according to the position of the teeth in the jaw. In this problem, we seek to automatically assign a label to each ROI that reflects its tooth-class, and hence we call this the teeth-labeling problem. Because teeth labels are important in construction of the dental chart, we also refer to this problem as automatic construction of the dental chart.

- b. **Micro and Macro decision-making:** our previous research [6][21] focused on the problem of tooth-to-tooth comparison using a single region of interest per tooth (*see* Appendix). In this problem we exploit two facets of teeth multiplicity in dental radiographic views: First, a tooth may appear in more than one dental radiographic film, hence when comparing different views of a subject/reference tooth-pair we are likely to obtain different match probabilities (match scores). One important question we address is on combining the match scores resulting from comparisons of the multiple views of a subject/reference tooth-pair and thus reaching a tooth-level match decision (micro decision). Second, as a dental record usually reflects presence of multiple teeth, it is logical to devise this aspect of multiplicity in making more reliable automatic case-level match decisions (macro decision) based on aggregation of the micro-decisions (tooth-level match decisions). We call this the micro and macro decision-making problem.
- c. **Performance Limits of Dental Features:** several features were studied in the context of automating the process of postmortem identification. The performance of a biometric system depends, among other factors, on the capability of the feature representation scheme to capture the intra-subject invariance and the inter-subject variance, ideally retained by the biometric signal. In this problem, we seek to determine bounds on the capability of two sets of dental features, the first uses teeth contours, and the second is appearance-based. We refer to this problem as performance limits of dental features.

1.5 Research Contributions

The contributions of this work are:

- A methodology for automatic construction of dental charts using low computational-cost, appearance-based features and string matching.
- A hierarchical fusion scheme for more robust dental identification that exploits multiple teeth and multiple views of a tooth in dental records.
- A study of performance limits of dental identification due to two sets of features.

1.6 Dissertation Organization

The remainder of this dissertation is organized as follows: in Chapter 2 we provide a background to the problem, we address and describe related work, then we report the literature survey we conducted, and finally we formally define the problems we research. We tackle the problem of automatic construction of dental charts in Chapter 3. In Chapter 4 we present a hierarchical fusion scheme that addresses micro and macro decision-making. In Chapter 5 we study the performance limits of dental identification as set by the dental features used. Finally in Chapter 6 we conclude the dissertation and outline opportunities for future work.

Chapter 2: Background and The Problem

Identification of deceased individuals based on dental characteristics is receiving increased attention, especially with the large volume of victims encountered in mass disasters. There have been several attempts for automating the dental identification problem; however, the achieved level of automation is relatively low.

In this chapter we provide a background to the problem of automated dental identification, we review the literature on the areas related to the three problems we tackle, and finally we present a formal problem statement.

2.1 Background

There have been several attempts to develop computer-aided postmortem identification systems. The most famous among these systems are *CAPMI* [5] and *WinID*[®] [22]. To the best of our knowledge none of the existing systems provide the desired level of automation, as they require a significant amount of human intervention. For example, in both *CAPMI* and *WinID*[®], feature extraction, coding, and image comparison are carried-out manually. Moreover, the dental codes used in these systems are predominantly based on dental work [5][22].

The Computer Assisted Post Mortem Identification system (*CAPMI*) was developed by the bioengineering branch of the US Army Institute of Dental Research. *CAPMI* is a computer software program that compares between dental codes extracted from *AM* and *PM* dental records. The program generates a prioritized list of candidates based on the number of matching dental characteristics. This list guides forensic odontologists to reference records that have potential similarity with subject records. The odontologist then completes the identification procedure by visual comparison of radiographs [5].

WinID[®] is a computer system that matches missing persons to unidentified persons using dental and anthropometric characteristics to rank possible matches (anthropometry is the study of human body measurements). Other information on physical appearances, pathological findings and anthropologic findings can also be added to the database of *WinID*[®]. The dental codes used in *WinID*[®] are extensions of those used in *CAPMI*.

In [6][21] we propose the use of learnable inherent dental image features for tooth-to-tooth image comparisons (*see* Appendix page 104). We treat the tooth-to-tooth matching problem as a binary classification problem for which we propose parametric models of class-conditional densities. We also propose an adaptive strategic searching technique and use it in conjunction with back propagation in order to estimate system parameters.

In [11], *Jain* and *Chen* propose a semi-automated system for human identification based on matching of teeth contours extracted from dental x-ray images. Their system follows three main steps for identification: radiograph segmentation, teeth contour extraction, and shape matching. For each radiograph, a human user initializes segmentation by specifying a pixel that belongs to the *gap valley* (an artificial curve that best separates the maxilla and mandible), then detection of the entire gap valley as well as teeth isolation are carried out using integral projection.

In [11], extraction of an isolated tooth contour is achieved in two steps, crown extraction and root extraction. Crown extraction is treated in a pixel classification approach along radial rays, while root extraction is considered an iterative neighborhood context analysis problem. Identification is completed by searching a database of contours for records with relatively small *Euclidean matching distance* to subject contours using a *quasi-affine* transformation model for contour alignment. In [12], *Chen* and *Jain*, propose a *directional SNAKE* approach for contour extraction of isolated teeth to mitigate excessive contour extraction errors of their approach in [11].

In [14] *Mahoor* and *Abdel-Mottaleb* present an approach for teeth classification in dental bitewing views using *Fourier* descriptors of teeth contours followed by a validation step based on string matching. In [15] *Jain* and *Chen* present an alternative approach for teeth classification also using teeth contours as features and Hidden Markov Models for registration of teeth to a human dental atlas.

2.2 Literature Review

In this section we present a review of the literature in areas that are closely related to the research problems we are interested in studying. In section 2.2.1 we review the literature on the attempts for automating the postmortem identification process, primarily highlighting the operation concepts of *CAPMI* and *WinID*[®]. In section 2.2.2 we briefly review the literature on image registration, as it is a tightly related area to the application in hand and provides ideas to explore for view normalization prior to teeth classification. In section 2.2.3 we review the concept of principal component analysis and look at its application in face detection. In section 2.2.4 we review important concepts and approaches in classifier fusion, which is a key concept for addressing the micro-to-macro decision-making problem. In section 2.2.5 we shed light on performance limits in biometrics, as basis for understanding possible venues for tackling the problem of determining performance limits of dental identification.

2.2.1 Postmortem Dental Identification

There have been several attempts to develop computer-aided postmortem identification systems. The most famous among these systems are *CAPMI* [5] and *WinID*[®] [22]. To the best of our knowledge none of the existing systems provide the desired level of automation, as they require a significant amount of human intervention. For example, in both *CAPMI* and *WinID*[®] feature extraction, coding, and image comparison are carried-out manually. Moreover, the dental codes used in these systems are predominantly based on dental work [2][5][22].

The Computer Assisted Post Mortem Identification system (*CAPMI*) was developed in 1983 by the US Army Institute of Dental Research (USAIDR) to improve the efficiency of forensic scientists by automating repetitive comparisons between object (reference) and key (subject) files. *CAPMI* produces ranked lists (starting with most-likely matches). Both non-dental and dental characteristics are incorporated in the comparison. *CAPMI* is a sorting tool, not an identification system, which is efficient and meaningful only when handling large numbers of cases [23].

The list guides forensic odontologists to the most probable *AM* matches to the submitted *PM* records, so that they can rapidly proceed with a positive identification by comparison of radiographs [23]. In CAPMI, data entry is performed either by a keyboard input or by an optical mark reader, the dental information associated with each tooth include:

- Missing tooth
- Anomalous condition
- Porcelain jacket crown
- Full coverage crown
- Non-metallic restoration ceramic or acrylic/metal
- Temporary restoration, $\frac{3}{4}$ crown
- Removable partial denture
- Unerupted tooth
- Mesial restoration
- Facial restoration
- Amalgam restoration
- Distal restoration
- Pointic root canal
- Deciduous tooth
- Cavity on tooth
- Occlusal restoration
- Lingual restoration
- Stainless steel crown
- Gold/cast metal restoration
- Treated tooth
- Virgin tooth

CAPMI uses up to 16-bits of information per tooth for dental data representation. For instance, if a tooth is present bit '0' is set otherwise it is reset. The other bits are used to represent the 5 surfaces of the tooth and to indicate if a crown is present or the tooth is root filled. The following textual notation is used to describe the characteristics of each tooth: A mesial cavity is entered as (*m*), a crown is entered as (*jmodbp*) to indicate that all five surfaces are restored (*modbp*) and a crown is present (*j*). The presence of a tooth is indicated by (.) and its absence is indicated by (x). Thus *UL6.mod* indicates that the Upper Left first permanent molar is present and has restorations in the mesial, occlusal and distal surfaces, while *LL3 x* indicates the Lower Left canine is missing. (*b: bucal, p: palatal*).

After entering the postmortem data for all the teeth, dental information of a subject is compared to the dental information of the candidates. The result is a list of the most likely matches, ordered descendingly from the most likely match to the least likely match. The comparison output is either: Match, Mismatch, or a Possible match. A Match occurs when the *AM* and *PM* dental data are exactly the same. A Mismatch occurs when *AM* dental condition is different from the *PM* dental condition given that it could have not evolved into the *PM* dental condition (e.g. a tooth with *AM* three restored surfaces can not evolve into a *PM* two restored surfaces tooth). A Possible match occurs when the *AM* dental condition is different from the *PM* dental condition but could have evolved into the *PM* condition (e.g. a tooth with *AM* two restored surfaces may evolve into a *PM* three restored surfaces tooth) [5].

While forensic scientists rely on teeth orientation, type of restorative materials, and radiographic appearance as basis for positive identification. These properties are not incorporated in CAPMI as “*testing has shown that incorporation of these additional data would only increase processing time while decreasing the power of the system due to mismatches induced by the subjectivity inherent in the recognition and identification of these entities*” [23]. Not only does CAPMI utilize dental codes that describe teeth conditions with regard to restored surfaces, but it also allows for incorporation of non-dental identifiers (e.g. gender, blood type, .. etc) into the records of [23].

In [24] *Williams et. al.* introduce another CAPMI ranking scheme that was motivated by an identification scenario they encountered in an air plan crash where remains were highly fragmented (112 PM records generated from fragments) and only a few AM records were present (6 records). The customary ranking scheme of {maximum matches, minimum mismatches, and possible matches} pushes the actual match down the list due to probable lack of “many” matches in a fragment (for example a short truly matched fragment would not get a good match score as a longer one with some matches and mismatches). Ranking fragmented remains using minimum mismatches first puts true matches up in the list compared to the maximum matches first scheme. In [25] *Lorton et. al.* study alternative manipulation schemes for records based on “*Master sort, Lower jaw sort, Upper jaw sort, Trimmed sort, Logical sort, Stratified sort, and Two-step sort*”.

In [26] *Friedman et. al.* study the dental characteristics (with dental characteristics being restored surfaces of teeth) of a large military population of 7030 soldiers and argue that despite the diminishing number of restored teeth among the younger generations in the US, the degree of selectivity of CAPMI as a sorting system is acceptable. Based on 363 simulations, they found that in 90% of the cases with two or more characteristics CAPMI gets a hit at the top of the candidate list, and that in ~96% of the cases CAPMI gets a hit within the top 4 candidates.

WinID[®] [22] is a dental computer system that matches missing persons to the unidentified persons using dental and anthropometric characteristics to rank possible matches. Other information about physical descriptors, and pathological and anthropologic findings can be fed to the *WinID*[®] database. The dental codes used in *WinID*[®] are extensions of the CAPMI

codes. Up to five primary codes may be specified and a dash (-) is placed between the primary and the secondary codes, however, most of the searches in WinID[®] utilize only the primary codes. Following are a listings of the primary and secondary codes used in WinID[®].

WinID[®] Primary Codes:

- **M: mesial** surface of tooth is restored.
- **D: distal** surface of tooth is restored.
- **L: lingual** surface of tooth is restored.
- **C: tooth** is fitted with a **crown**.
- **V: non-restored tooth – virgin**.
- **J: tooth** is **missing postmortem** or the clinical crown of the tooth is not present for examination. Also used for avulsed tooth. The root or an open socket is present, but no other information is available.
- **/: no information** about tooth is available.
- **O: occlusal** surface of posterior tooth is restored.
- **F: facial** surface of tooth is restored.
- **I: incisal** edge of anterior tooth is restored.
- **U: tooth** is **unerupted**.
- **X: tooth** is missing – **extracted**.

WinID[®] Secondary Codes:

- **B: tooth** is **deciduous**.
- **G: gold** restoration.
- **R: root** canal filled.
- **A: anomaly** is associated with this tooth. Specifics of the anomaly may be detailed in the comments section.
- **N: non-precious** filling or crown material. Includes stainless steel.
- **P: pontic**. Primary code must be X to indicate missing tooth.
- **Q: three quarter** crown. Primary code must be C to indicate crown.
- **T: denture tooth**. Primary code must be X to indicate missing tooth.
- **Z: temporary** filling material. Also indicates gross **caries**.
- **E: resin** filling material.
- **H: porcelain**.
- **S: silver** amalgam.

In addition to the textual representation of the dental codes, WinID[®] also provides a visual representation of the dental codes in the form of a dental chart. Figure 3 shows an example of a dental chart produced by WinID[®]. Each tooth is assigned a two-digit number; the most significant digit indicates the jaw segment in which the tooth lies, while the second digit indicates the position of the tooth in its jaw segment. The right side of the upper jaw is assigned '1', the left side of the upper jaw is assigned '2', the left side of the lower jaw is assigned '3', and right side of the lower jaw is assigned '4'. Front teeth are assigned position number '1' in each segment, and numbering proceeds through '8' which is assigned to the wisdom tooth in each segment.

In [27] McGivney highlights the commonalities and differences between CAPMI and WinID[®]; both systems use the same comparison algorithm and are equally capable of ranking possible identifications by least number of dental mis-matches [23] or most dental hits.

However, only WinID[®] -but not CAPMI- is capable of generating a raked list of “non-dental” identifier matches as well as a list of most restoration hits (which precludes virgin teeth and missing teeth). WinID[®] also features generation of odontograms (dental charts) and is capable of displaying images, which are two very distinguishing features of WinID[®] over CAPMI.

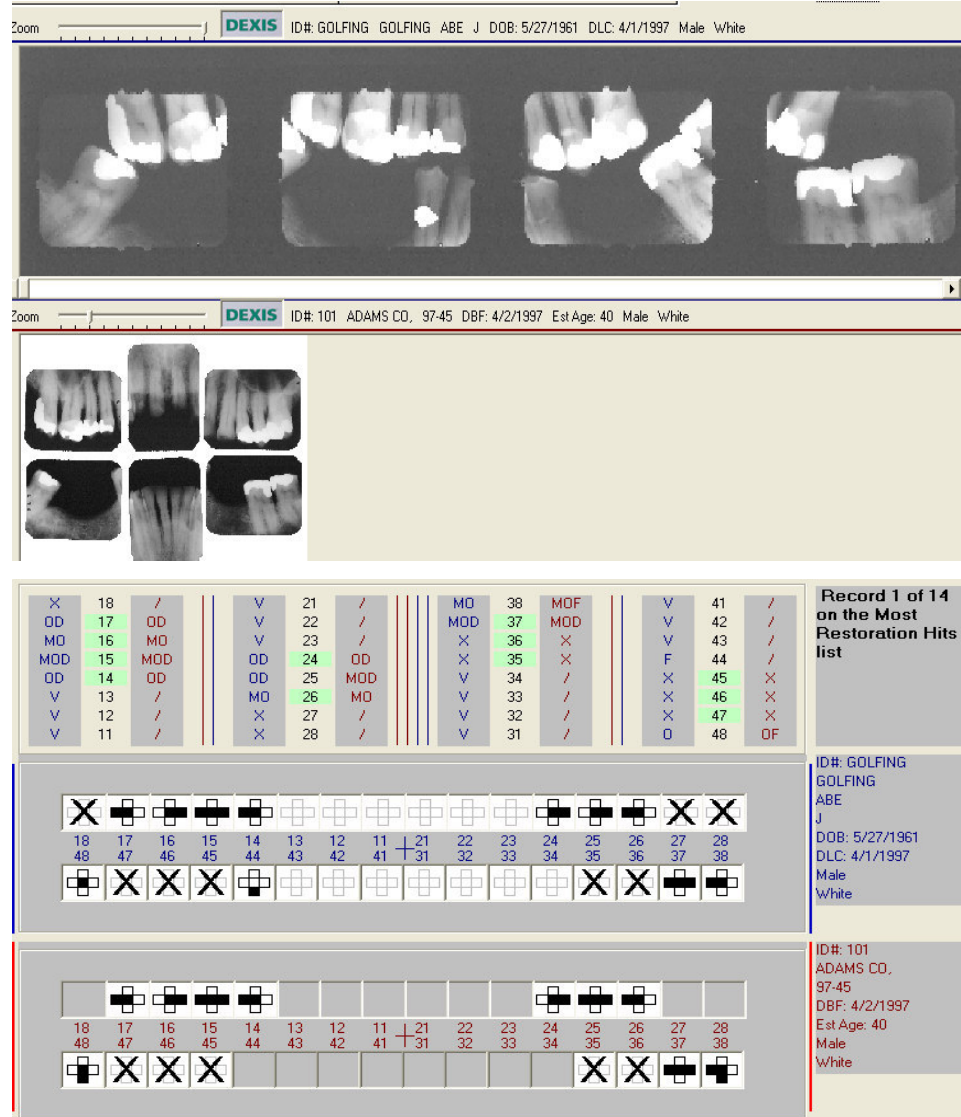


Figure 3: Example of using dental charts for identification in WinID[®].

In [28] *Lewis*, and *Leventhal* hold a comparison between CAPMI, WinID[®], and a *Locator System* algorithm that they proposed. The main difference between the algorithm they proposed and those used by CAPMI, and WinID[®] is that the *Locator System* algorithm relies

on initial screening and categorization of AM and PM files into 6 categories based on the characteristics of dental work, i.e. whether there is any restoration, the types of restored surfaces, presence of crowns and roots, missing teeth, .. etc. *Lewis*, and *Leventhal* assert that with few or no restorations, all three systems perform equally poor.

In [29] *Bowers* and *Cornwell* use operator controlled digital image processing and analysis techniques (provided by a commercial software Adobe[®] Photoshop[®] 5.0) to visually compare PM records of two cases to their candidate AM records. For the first AM/PM pair, they establish a common baseline between a tooth-pair by adjusting the orientation of films such that the cemento-enamel junction (CEJ) of the tooth-pair appears horizontal. Thereof they resize the PM tooth such that its CEJ dimension agrees with that of the AM tooth, and hence they measure two age-independent features: the root furcation heights and the distal root divergence angle. By observing significant difference in these measures, *Bowers* and *Cornwell* reject the hypothesis that the first PM case matches its candidate AM case. With the second AM/PM pair they map a smoothed restoration region from the AM tooth to the PM tooth using an alignment transform and conclude that the PM record matches the AM record based on their visual comparison of shapes of aligned restorations.

In [3], *MacLean et. al.* conduct a controlled experiment to validate to the adequacy of human identification using bitewing dental radiographs. They formed two groups of bitewing radiographs: a matched group and an unmatched group. The former group contained 140 pairs of radiographs acquired over diverse time periods (1-15 years) from 83 individuals, thus simulating postmortem records. The unmatched group had 140 unmatched pairs acquired from different individuals. To count for the diminishing number of restorations in contemporary generations, they used radiographs of individuals that either had no dental restorations or whose dental restorations are do not provide significant clues for identification. Three observers worked independently on bitewing comparisons (two experienced dentist and a senior year dental student, one of the experts had considerable experience in forensic dentistry), accuracy of decisions was 93%, with less number of false positives achieved by the expert. *MacLean et. al.* assert that false positives are the crucial errors as they may result removal of a wrong record form the from the pool, while false negatives may be mitigated by retesting. Their study concludes that high accuracy of human identification can be achieved even with a “*substantial time lapse*” between AM and PM

radiographs, bitewing radiographs are a valid tool for positive identification based not only on metallic restorations, but also on anatomic characteristics.

In [30] *Soomer et. al.* instrument a statistical testing approach to study the human factors due to training, experience and work sector that affect the accuracy of radiographic identification. Their study involved 40 participants from 19 countries that were either affiliated with academia, governments, or private sectors. Moreover these participants were subject to different levels of training assessed by degrees, fellowships, participation in forensic odontology training courses, and active membership in professional forensic societies. Participants also belonged to different experience categories assessed based on years of experience in the forensic odontology field, number of identification cases worked out annually, participation in mass disaster victim identifications, and whether they maintain an active affiliation with a forensic institute (or team). Each participant was presented 9 AM records (containing bitewing and apical radiographs) and 9 PM real (not simulated) records of panoramic radiographs. Participants were asked to compare the AM and the PM cases and report matching results using the categories and terminologies of the American Board of Forensic Odontologists (ABFO), i.e. positive identification, possible identification, insufficient evidence, and exclusion. These case-pairs varied in the number of “*match points*” and accordingly were not equally easily identified. The study finds that academicians and government fellows outperformed private sector practitioners, which is highly likely due to the formers achieving and maintaining high levels of training.

While comparison of dental radiographs facilitates PM identification of victims, Bite marks, which are impressions of dentition on skin of victims, server the purpose of identifying criminals usually involved in physical assaults [31][32].

2.2.2 Image Registration

Image alignment (also referred to as image registration) is the process of transforming an image of an object (the *subject* or *query* image), and measuring its resemblance either to another image of the same object, or to an image of another object (the *reference* image) [33]. *Brown* [34] identifies four facets for the image matching problem: the representation of image information - or choice of a feature space; the class of possible transformations - or choice of a search space; the choice of a search strategy for convergence to an optimal transformation; and the choice of a measure of closeness to an optimal solution - or a similarity metric.

Image features range between highly informative low-level features such as raw pixel intensities and detail suppressing high-level features like pattern grammars. Examples of features commonly used for image representation are: edge points (including contours), spectral contents of images, salient points derived from contours and surfaces (e.g. points of maximum curvature, centers of gravity), and statistical features (like moments) [34][35][36].

Transformations may be categorized into geometric transformations and photometric transformations. While the former restore the geometric properties of imaged objects, the latter restore the intensities of corresponding pixels of the subject and reference images. Examples of geometric transformations are: rigid body (including rotation, scaling, and translation), quasi-affine (which accommodates change in aspect ratios), affine (which accommodates shear), projective, perspective, and polynomial [34][35][36].

Choosing a suitable search strategy is largely influenced by characteristics of the search space. If the search space is bounded by linear inequalities, for instance, then linear programming becomes the most suitable search strategy [34]. Relaxation, dynamic programming, and the generalized Hough transform are among other optimization strategies used in image matching [34].

Choice of a suitable similarity metric is usually tied to the feature space selection, as their combination should ultimately ignore uncorrectable distortions [34]. Cross-correlation functions, correlation coefficient, Euclidean distance, and difference images are among many similarity metrics that are usually used to quantify the degree of closeness of features in one image to those in the other under a candidate transformation [34][35].

There are also other aspects used for categorization of image matching problems, which include: image dimensionality, modality of images, whether reference images are templates, whether objective is to register object images taken from different viewpoints, and the desired (permissible) level of automation (human intervention) [34][36].

Image matching methods may be classified according to the choices made in addressing these facets of the problem. For example there are two root classes of feature spaces, namely intensity features and geometric features. Accordingly content-based image matching methods may broadly be classified into: intensity-based methods where color and/or texture information serve as basis for matching, and geometry-based methods where various shape information may be used for registration [33].

We must also differentiate between three aspects of image matching: transformation, variations, and computation. Transformation refers to mapping of pixels in one image to new locations in another. Variations refer to the differences in the pixel values and locations between images, which include distortions that corrupt true measurements. And computation refers to the set of calculations used to estimate the parameters of a transformation [33][34]. Based on the neighborhood used in tackling each of these aspects, transformations, variations and computation may be designated as either local or global [33][34].

In radiography, affine geometry perfectly describes imaging with infinite distance between tube and subject, which causes the beams to be parallel [36]. On the other hand, projective geometry perfectly describes X-ray imaging with an infinitely small focal spot size, which is equal to a point source [36]. In [37], *Ostuni et. al.* present an experimental study of the geometric nature of dental radiographs, they conclude that dental radiograph registration should generally assume projective geometry. However, when an object is close to being perpendicular to the optical axis of the radiography apparatus, registration assuming affine geometry is as effective as that assuming projective geometry [37].

In [38] *Yoon* presents an automated method for digital subtraction radiography of dental radiographic images. Yoon's approach for alignment uses 1500 or more edge pixels from each dental radiographic image, provided that images are taken for the same anatomical region of a known patient. One-to-one correspondences between edge features in both images are established using a nearest neighbor criterion. Edge features in the subject

radiograph are aligned with those in the reference radiograph by performing an affine transformation. To determine affine transformation parameters, Yoon iteratively seeks solutions to minimize the Euclidean distance (computed efficiently using distance maps) between reference edge pixels and their nearest neighbors of subject edge pixels (under iterative transformation).

In [39] *Zacharaki et. al.* propose a multiresolution approach for digital subtraction radiography in which they vary the complexity of the geometric transformation model across the multiresolution levels of the subject and reference images. At the coarser levels, they employ affine transforms, at the image level they rely on projective transforms so that they ultimately achieve a high accuracy registration. The similarity measure they use is the normalized correlation coefficient, with features being pixel intensities.

In [40] *Samadzadegan et. al.* propose a method for matching different types of radiographic films based on a multiresolution fuzzy reasoning strategy for determining corresponding pixels in the subject and reference radiographs. Their fuzzy reasoning strategy takes into account geometric constraints, radiometric similarities, and textural differences, all of which they believe to influence the human decisions of establishing correspondences between pixels.

2.2.3 Principal Component Analysis

Often in pattern recognition tasks, one is faced with features of high dimensionality [41]. Not only is processing of high-dimensionality features computationally expensive, but also it is likely to raise the error rate in classification and decision-making, primarily due to the *curse of dimensionality* problem [42][43]. It is therefore logical to reduce the features dimensionality in a way that does not significantly affect the objective of the pattern recognition task.

There are several approaches for reducing the dimensionality of features, some approaches are based on feature selection [44] and some others are based on mapping (or projection) of features onto subspaces (of the total feature space).

We focus on a subclass of feature mappings, namely *linear* mappings. Common interest in linear mappings is primarily due to their feasibility from a practical point of view and also due to their tractability from a mathematical point of view. In this section we review the concept of principal component analysis as an important technique for feature reduction, and we briefly overview three appearance-based approaches used in the area of face detection and recognition.

Principal Component Analysis (PCA) is a technique used for reducing data dimensionality while preserving the most significant variances in a given dataset. PCA is thus viewed as a *data representation scheme* that seeks to project a given dataset onto a subspace that “*best represents the data in a least-squares sense*” [41].

Given a dataset of N d -dimensional samples (i.e. $\{\underline{x}_i: i = 1 \dots N, \underline{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{id})\}$), one may inquire about the “*best*” way to represent all samples by a single point (say \underline{x}_o) in the d -dimensional space. The word *best* in the previous statement automatically calls for defining a criterion for evaluation. If we use the square-error criterion function $J_o(\underline{x}_o) = \sum_{\langle i \rangle} \|\underline{x}_i - \underline{x}_o\|^2$, it is easy to show (by differentiation along the d -dimensions) that in the least square-error sense, the best \underline{x}_o -as one would intuitively expect - is indeed the sample mean

$$\underline{x}_o = \underline{\mu} = \frac{1}{N} \sum_{\langle i \rangle} \underline{x}_i .$$

The sample mean (a point in the d -dimensional space) is considered a “zero-dimensional representation” of the sample [41]. It is obvious, however, that the mean does not capture any variability in a given dataset. So, in order to obtain a better representation of the dataset, we start the establishment of a subspace by identifying its first direction. Hence, we seek to represent the different samples as: $\underline{x}_i = \underline{\mu} + a_i \underline{u}$, where \underline{u} is a unit vector (defining a dimension of the sought subspace) and a_i is a real valued scalar that accounts for the projection of \underline{x}_i on the subspace along its first dimension (defined by \underline{u}).

The best direction (\underline{u}) for such a representation in a least square-error sense is obtained by minimizing the criterion function:

$$J_1(a_1, \dots, a_i, \dots, a_N, \underline{u}) = \sum_{\langle i \rangle} \|\underline{x}_i - (\underline{\mu} + a_i \underline{u})\|^2 = \sum_{\langle i \rangle} a_i^2 - 2 \sum_{\langle i \rangle} a_i \underline{u}^T (\underline{x}_i - \underline{\mu}) + \sum_{\langle i \rangle} \|\underline{x}_i - \underline{\mu}\|^2.$$

By equating the partial derivatives of J_1 w.r.t a_i 's to zeros, we thus obtain: $a_i = \underline{u}^T (\underline{x}_i - \underline{\mu})$.

Now to identify the direction of \underline{u} we rewrite $J_1(\underline{u})$ as:

$$J_1(\underline{u}) = - \sum_{\langle i \rangle} \underline{u}^T (\underline{x}_i - \underline{\mu}) (\underline{x}_i - \underline{\mu})^T \underline{u} + \sum_{\langle i \rangle} \|\underline{x}_i - \underline{\mu}\|^2 = - \underline{u}^T S \underline{u} + \sum_{\langle i \rangle} \|\underline{x}_i - \underline{\mu}\|^2, \text{ where } S \text{ is the scatter matrix of the dataset given by } S = \sum_{\langle i \rangle} (\underline{x}_i - \underline{\mu}) (\underline{x}_i - \underline{\mu})^T, \text{ it obvious that for } \underline{u} \text{ to minimize } J_1(\underline{u})$$

it has to equivalently maximize the term $\underline{u}^T S \underline{u}$ with the constraint of $\|\underline{u}\|=1$.

Thus, the problem reduces to a constrained maximization that can be solved using the *Lagrangian* formulation: $\max_{\langle \underline{u} \rangle} \{L(\underline{u}) = \underline{u}^T S \underline{u} - \lambda (\underline{u}^T \underline{u} - 1)\}$ resulting the *eigenvalue* problem:

$(S - \lambda I)\underline{u} = 0$. Therefore, maximizing $L(\underline{u})$ calls for choosing the *eigenvector* \underline{u} corresponding to the largest *eigenvalue*.

In summary, the best one-dimensional projection of a dataset (in a least squared-error sense) is obtained using a line parallel to the eigenvector corresponding to the largest eigenvalue and passing through the sample mean.

Using similar analysis, the problem of projection using linear mapping can be extended to M dimensions ($M < d$), in which case:

$\underline{x}_i = \underline{\mu} + \sum_{j=1}^M a_{ij} \underline{u}_j$, and $J_M(\underline{u}_1, \dots, \underline{u}_M) = \sum_{\langle i \rangle} \left\| \underline{x}_i - \left(\underline{\mu} + \sum_{j=1}^M a_{ij} \underline{u}_j \right) \right\|^2$, where minimization of $J_M(\{\underline{u}\})$

is achieved by using the eigenvectors corresponding to the M largest eigenvalues [41]. Because the scatter matrix S is real and symmetric, eigenvectors are orthogonal and constitute the basis defining the sought M -dimensional subspace.

The M -coefficients corresponding to the projections an N -dimensional data point \underline{x} on the M -dimensional subspace are called the principal components of \underline{x} . And in effect PCA using M basis-vectors restrains attention to the M directions of greatest scatter in the dataset.

2.2.3.1 EigenFaces, FisherFaces, and LaplacianFaces

In [45] and [46] *Turk* and *Pentland* introduce a PCA-based approach for face detection and recognition, whose motivation is attributed to earlier work by *L. Sirovich* and *M. Kirby* on low-dimension representation of human face pictures for efficient picture storage (published in 1987). *Turk* and *Pentland* build their analysis on two main assumptions: (i) faces are captured in upright views, and (ii) faces can be represented by a small set of 2D characteristic views. To establish the set of characteristic views, they use a training set of faces without mandating a specific set of isolated face features (e.g. mouth, nose, and eyes).

The main idea of face recognition using *eigenfaces* is to define a face space using the eigenvectors captured from the training set of faces. These eigenvectors encode the information that best describe the variation in the exemplary set of faces. Then during the recognition phase, face images are then projected onto the face space (defined by the eigenvectors) to obtain coefficients that are believed to be distinctive for different faces [45][46].

Turk and *Pentland* further propose that only few eigenfaces are sufficient to constitute the face space, thus an N^2 -dimensional image space is reduced to an M -dimensional subspace defined by the eigenvectors corresponding to the largest M eigenvalues.

In [45], *Turk* and *Pentland* propose an efficient scheme to determine the eigenvectors of the large $N^2 \times N^2$ scatter matrix, by linearly relating the eigenvectors of the scatter matrix with those of a simpler $M \times M$ matrix formed from the set of training face images, thus greatly reducing the computational cost of the training phase.

The eigenfaces approach for face detection/recognition is reported to exhibit low performance under pose and illumination variations [47]. We also recall that the objective function used in the deriving of the PCA method is such that the directions along which the training dataset exhibit maximum variations are the preferred directions [41]. In [47], *Belhumeur et. al.* provide a very interesting 2D example where PCA fails to fulfill a classification task, this is not due to a flaw in the PCA theory, it is rather a consequence of its inappropriate application in a particular problem. PCA is primarily a tool for efficient representation.

In contrast with the PCA, Multiple Discriminant Analysis (MDA) offer linear mappings that best separate data from different classes in a least square-error sense [41]. *Belhumeur et. al.* [47] propose using MDA to create a facespace using *Fisherfaces*.

In [48], *He et. al.* establish that both PCA and MDA are geared towards subspaces that preserve the global structure of objects, they constitute that in many real-world problems the local structure is more important than the global structure. Hence, they propose a Locality Preserving Projection (LPP) approach leading to *Laplacianfaces* in the case of face recognition.

2.2.4 Classifier Fusion

Traditionally, researchers and practitioners tackled the problem of pattern recognition by adopting a single classification scheme that experimentally proves to possess the best performance amongst other possible classification schemes [49][50]. Practically, it has been observed that misclassified patterns under each of the classification schemes do not necessarily overlap and thus the different classification schemes may provide complementary information about the patterns to be classified [49]. This observation motivated research on combining classifiers (also known as classifier fusion/integration). Also with the large dimensionality of features, it becomes desirable to divide the classification task between a number of classifiers, each of which works on a subset of the feature dimensions, and to integrate the outcome of the classifiers as opposed to training a single classifier using all dimensions of the feature set [50]. Not only does classifier fusion provide faster training and reduced classifier complexity, but it also improves classification performance and robustness [50][51].

Information fusion in the context of classification applications may take place either prior to applying any classification algorithm (*pre-mapping* or *pre-classification* fusion), after application of multiple classification schemes (*post-mapping* or *post-classification* fusion), or while mapping features to the decision space (*midst-mapping* fusion) [52][53].

In [54] *Ross* and *Jain* name several advantages for fusion in biometric systems, mainly: using multiple pieces of evidence to achieve more reliable decisions compared to systems based on a single *noisy* biometric, overcoming the non-universality of some biometrics in some user groups, and offering opportunities for a more robust *vitality* detection. They identify three possible levels¹ for fusion in multi-modal biometric systems (and in our opinion in other classification problems too):

- (i) Feature level fusion; by combining a set of small feature vectors into a larger dimension feature vector (either by concatenation or weighted summation [52],

¹ [53] contains an enriched hierarchy of information fusion approaches.

- (ii) Matching score level fusion (or posterior probability level fusion) by combining match scores of the different matchers (or the posterior probabilities of the different classifiers) into a single match score, and
- (iii) Decision level fusion by combining classifier decisions of multiple classifiers.

Ross and Jain [54] also identify six variants for fusion in the context of biometric systems specifically:

- (i) Multiple representations of a single biometric, with possible fusion at either the feature level or the matching score level,
- (ii) Multiple matchers and a single biometric representation, with fusion at the matching score level,
- (iii) Multiple biometric identifiers, with fusion at the matching score level after mapping different scores to a common domain for normalization,
- (iv) Multiple versions (or templates) of a reference biometric, with fusion at the matching score level to integrate scores of matching of a subject against all templates of a reference,
- (v) Multiple versions (or impressions) of a subject biometric, where the multiple acquisitions of the subject biometric are matched to a single reference biometric and fused to produce a single match score, and
- (vi) Multiple distinct instances of a biometric, e.g. multiple fingerprints, or left and right eye irises, are matched to their corresponding instances in the reference's biometric and matching scores are fused to produce a single score.

Various schemes may be used for score normalization when fusion is carried out at the matching score level, examples of which are: the min-max normalization, the decimal scaling, the z-score normalization, the median/median-absolute-deviation, the double sigmoid normalization, the tanh-estimator normalization, and the biweight estimator normalization [52].

In [49] and [50] *Kittler et. al.* present a unified theoretical framework for classifier fusion using distinct feature representations. They formulate the problem as:

- Given:

- m -classes
- R -Classifiers
- i^{th} classifier uses \mathbf{X}_i feature vector.
- Class conditional pdfs are : $p(\mathbf{X}_i | \omega_k)$; $i = 1, \dots, R$ and $k = 1, \dots, m$
- Prior $p(\omega_k)$
- Training “models” are uniquely assigned to classes.

- The Problem:

- Find the class of novel pattern Z whose R -feature vectors (descriptors) are: $\{\mathbf{X}_i, i = 1, \dots, R\}$ using *Bayes* decision rule:

$$Z \rightarrow \omega_j | p(\omega_j | \mathbf{X}_1, \dots, \mathbf{X}_R) = \max_{\langle i \rangle} p(\omega_i | \mathbf{X}_1, \dots, \mathbf{X}_R)$$

Computing $p(\omega_i | \mathbf{X}_1, \dots, \mathbf{X}_R)$ requires knowledge of the joint probability densities:

$$p(\omega_k | \mathbf{X}_1 \dots \mathbf{X}_R) = \frac{p(\mathbf{X}_1 \dots \mathbf{X}_R | \omega_k) p(\omega_k)}{\sum_{i=1}^m p(\mathbf{X}_1 \dots \mathbf{X}_R | \omega_i) p(\omega_i)} \quad (1)$$

The denominator of (1) is a *normalization* term, which is constant irrespective of the class ω_k and accordingly decision-making depends only on the numerator of (1). *Kittler et. al.* [49][50] derive the following fusion rules:

A) The Product rule: assuming $(\mathbf{X}_1, \dots, \mathbf{X}_R)$ are independent, then $p(\mathbf{X}_1, \dots, \mathbf{X}_R | \omega_k) =$

$$\prod_{i=1}^R p(\mathbf{X}_i | \omega_k). \text{ Thus, } p(\omega_k | \mathbf{X}_1 \dots \mathbf{X}_R) = \frac{p(\omega_k) \prod_{i=1}^R p(\mathbf{X}_i | \omega_k)}{\sum_{j=1}^m p(\omega_j) \prod_{i=1}^R p(\mathbf{X}_i | \omega_j)} \text{ and accordingly Bayes decision}$$

rule reads:

$$Z \rightarrow \omega_j | p(\omega_j) \prod_{i=1}^R p(\mathbf{X}_i | \omega_j) = \max_{\langle k \rangle} p(\omega_k) \prod_{i=1}^R p(\mathbf{X}_i | \omega_k) \quad (2)$$

Using inversion the product decision rule reads:

$$Z_{\pi} \rightarrow \omega_j | (p(\omega_j))^{-(R-1)} \prod_{i=1}^R p(\omega_j | \mathbf{X}_i) = \max_{\langle k \rangle} (p(\omega_k))^{-(R-1)} \prod_{i=1}^R p(\omega_k | \mathbf{X}_i) \quad (3)$$

It is worth noting that any diminishingly small posterior $p(\omega_k | \mathbf{X}_i)$ pulls the combined classifier posterior $p(\omega_k | \mathbf{X}_1, \dots, \mathbf{X}_R)$ close to zero, and accordingly rules out ω_k . Moreover, if this was the case with all ω_k classes, all posteriors would be very close to zero, which does not constitute a desirable case.

B) The Sum rule: in addition to independence between dimensions of the feature vector, assume further that posteriors slightly deviate from priors. Thus, $p(\omega_k | \mathbf{X}_i) = p(\omega_k) (1 + \delta_{ki})$, substituting $p(\omega_k | \mathbf{X}_i)$ back in the product decision rule and neglecting 2nd and higher order terms, we obtain the sum rule which reads:

$$Z_{\Sigma} \rightarrow \omega_j | (1-R) p(\omega_j) + \sum_{i=1}^R p(\omega_j | \mathbf{X}_i) = \max_{\langle k \rangle} \{ (1-R) p(\omega_k) + \sum_{i=1}^R p(\omega_k | \mathbf{X}_i) \} \quad (4)$$

By observing that: $\prod_{i=1}^R p(\omega_j | \mathbf{X}_i) \leq \min_{\langle i \rangle} p(\omega_j | \mathbf{X}_i) \leq \frac{1}{R} \sum_{i=1}^R p(\omega_j | \mathbf{X}_i) \leq \max_{\langle i \rangle} p(\omega_j | \mathbf{X}_i)$ other classification strategies are derived.

C) The Max rule: starting with the sum rule, if the sum of posteriors is approximated using the maximum posteriors, the decision rule becomes:

$$Z_{max} \rightarrow \omega_j | (1-R) p(\omega_j) + R \max_{\langle i \rangle} p(\omega_j | \mathbf{X}_i) = \max_{\langle k \rangle} \{ (1-R) p(\omega_k) + R \max_{\langle i \rangle} p(\omega_k | \mathbf{X}_i) \} \quad (5)$$

The max rule with the assumption of equal priors becomes:

$$Z_{max} \rightarrow \omega_j | \max_{\langle i \rangle} p(\omega_j | \mathbf{X}_i) = \max_{\langle k \rangle} \{ \max_{\langle i \rangle} p(\omega_k | \mathbf{X}_i) \} \quad (6)$$

D) The Min rule: starting with the product rule, and considering the supremum of the posteriors product, the decision rule becomes:

$$Z_{min} \rightarrow \omega_j | (p(\omega_j))^{-(R-1)} + R \min_{\langle i \rangle} p(\omega_j | \mathbf{X}_i) = \max_{\langle k \rangle} \{ (p(\omega_k))^{-(R-1)} + R \min_{\langle i \rangle} p(\omega_k | \mathbf{X}_i) \} \quad (7)$$

The min rule with the assumption of equal priors becomes:

$$Z_{min} \rightarrow \omega_j | \min_{\langle i \rangle} p(\omega_j | \mathbf{X}_i) = \max_{\langle k \rangle} \{ \min_{\langle i \rangle} p(\omega_k | \mathbf{X}_i) \} \quad (8)$$

E) The Median rule: under equal priors assumption, the sum decision rule may be viewed as an average rule, i.e:

$$Z_{average} \rightarrow \omega_j \mid \frac{1}{R} \sum_{i=1}^R p(\omega_j \mid \mathbf{X}_i) = \max_{\langle k \rangle} \left\{ \frac{1}{R} \sum_{i=1}^R p(\omega_k \mid \mathbf{X}_i) \right\} \quad (9)$$

The average decision rule may lead to incorrect decisions because the average is a sensitive measure to outliers. By adopting the median of the posteriors, the corresponding decision rule becomes:

$$Z_{median} \rightarrow \omega_j \mid \text{med}_{\langle i \rangle} p(\omega_j \mid \mathbf{X}_i) = \max_{\langle k \rangle} \left\{ \text{med}_{\langle i \rangle} p(\omega_k \mid \mathbf{X}_i) \right\} \quad (10)$$

F) The Majority vote rule: by hardening the posteriors in the sum rule (similar to winner-take-all) according to $p(\omega_k \mid \mathbf{X}_i) = \Delta_{ki} = \begin{cases} 1 & \text{if } p(\omega_k \mid \mathbf{X}_i) = \max_{\langle j \rangle} p(\omega_j \mid \mathbf{X}_i) \\ 0 & \text{otherwise} \end{cases}$. The resulting

decision rule becomes:

$$Z_{Majority} \rightarrow \omega_j \mid \sum_{i=1}^R \Delta_{ji} = \max_{\langle k \rangle} \left\{ \sum_{i=1}^R \Delta_{ki} \right\} \quad (11)$$

The sum $\sum_{i=1}^R \Delta_{ki}$ is a count of votes received from all R classifiers on the hypothesis that the observation vector \mathbf{X}_i belongs to class k . The class that receives the highest vote count is called out.

In [49][50][55], Kittler *et. al* offer a sensitivity analysis of error decision-making due to error in estimating posterior probabilities, specifically with the product rule and the sum rule. As decision-making utilizes the available distributions, which are rather estimates of the posteriors $\hat{p}(\omega_k \mid \mathbf{X}_i)$ that deviate from the true posteriors according to $\hat{p}(\omega_k \mid \mathbf{X}_i) = p(\omega_k \mid \mathbf{X}_i) + e_{ki}$, where $p(\omega_k \mid \mathbf{X}_i)$ is the true posterior and e_{ki} is the estimation error. With the assumptions that $e_{ki} \ll p(\omega_k \mid \mathbf{X}_i)$ and $p(\omega_k \mid \mathbf{X}_i) \neq 0$, Kittler *et. al* show that the

error-free combined-classifier decisions will be subject to an error factor ϵ ,

$$\epsilon = \begin{cases} 1 + \frac{\sum_{\langle i \rangle} e_{ki}}{\sum_{\langle i \rangle} p(\omega_k | \mathbf{X}_i)} & \text{product rule} \\ 1 + \frac{\sum_{\langle i \rangle} e_{ki}}{\sum_{\langle i \rangle} p(\omega_k | \mathbf{X}_i)} & \text{sum rule} \end{cases} .$$

As $p(\omega_k | \mathbf{X}_i) < 1$ (for most classifiers), the estimation error e_{ki} is amplified in case of the product combination rule, on the other hand estimation error is likely to get damped in case of the sum rule as the sum of posteriors for the most probable class ($\sum_{i=1}^R p(\omega_j | \mathbf{X}_i)$) is likely to be greater than one. It is interesting to observe that while the sum rule is driven under more stringent assumptions than the product rule, the sum rule for classifier combination is more robust than the product rule.

Jain et. al. [56] study the problem of fusion in matching multiple fingerprints as well as the problem of fusion in matching multiple impressions of a fingerprint. If the matching system produces a *micro* decision (or a hardened score) for each finger (or finger impression) fusion is only possible at an abstract level using the majority rule. If, however, the multiple match scores are made available then using strong statistical independence arguments, the product rule for classifier fusion under *Neyman-Person* (not *Bayesian*) decision-making is employed for obtaining reliability-improved decisions.

2.2.5 Performance Limits in Biometrics

Performance of biometrics-based personal identification systems is primarily limited by *accuracy* aspects (or *error* rates) [57]. To meet identification accuracy requirements, a biometric system should accommodate two types of variability: (i) *within-subject* (or *intra-class*) variability, and (ii) *between-subject* (or *inter-class*) variability [57][58]. It is generally desirable that a biometric identifier minimizes the first type of variability, thus minimizing the *false-reject rate* and maximizes the second type of variability, thus minimizing the *false accept rate* [59].

Jain and Pankanti [57] categorize the factors limiting the accuracy of a biometric identification system into: (i) “*information-limited behavior*”, (ii) “*representation-limited behavior*”, and (iii) “*invariance-limited behavior*”. The first class of factors pertains to the biometric signal acquisition process, where the biometric phenomenon, the environment and the sensor characteristics set a *lower error bound* (or a *higher performance bound*). The second class of accuracy-limiting factors is due to imperfections of the feature representation scheme, which may not retain all the useful information (intra-class invariance and inter-class variations) thus *raises* the error bound above (or *lowers* the performance bound below) that due to inherent information-limitation of the biometric signal. The last class of factors tends to further *raise* the error bound by failing to perfectly absorb the *intra-class invariance* of all subjects.

In [57][60] Jain *et. al.* raise a fundamental question in biometrics: “*what is the inherent discriminable information available in the input signal ?*” and argue that this question -as such- is “*ill-formulated*” as for instance it is not always possible to define the ideal response of a matcher to a *noisy* input signal.

In [59] Daugman classifies biometric features into *genotypic* features (e.g. *DNA*) and *phenotypic* features (e.g. *fingerprints*, and *iris*). While the former are genetically completely determined, the latter are not as they manifest as a result of the interaction between genetic constitution, development process and the environment. Daugman describes *desirable* biometric features as ones that possess “*little genetic penetrance*”, high complexity (*randomness*), and lifetime stability.

Daugman [59] empirically studied iris individuality by analyzing a large database of *IrisCodes*[®] (each *IrisCode*[®] is 2^{11} bits that encodes texture features extracted from an iris image), he found the mean *Hamming* distance between the *IrisCodes*[®] database vectors to be 0.498, with a standard deviation of 0.0306. Thus, he estimated that the number of degrees of freedom (*dof*) in the *IrisCode*[®] is 266. So, out of the 2,048 bits, only 266 are independent. According to *Daugman*'s study, the probability that two *IrisCodes*[®] disagree in m bits is modeled by $\binom{266}{m} \left(\frac{1}{2}\right)^{266}$. This suggests that it is highly improbable for two *IrisCodes*[®] to coincidentally agree without belonging to the same person. At the same time it is also highly improbable that *IrisCodes*[®] of two different persons agree in a significant number of *IrisCode*[®] bits.

Daugman [59] also conducted a similar study on *genetically identical* iris pairs using a database of left eye and right eye irises. The conclusion of that study is that the features extracted from iris texture are phenotypic. *Daugman*'s studies in [59], represent an approach for addressing the expressive (representational) power of a biometric feature.

In [59] *Daugman* concludes that “*Biometrics that lack the so many degrees-of-freedom and therefore lack this huge combinatorial property would be limited to use in one-to-one verification mode, or to searching only rather small databases*”.

Daugman [58][59], *Jain* and *Pankanti* [57] agree that in addition to the *degrees-of-freedom* (as a measure of the features' spread) needed for *discriminability*, significant *separation* between genuine and imposter biometric features' distributions is also required to capture the *invariance*. One of the measures used to quantify separation between genuine and imposter distributions is the decidability index $d' = \frac{|\mu_G - \mu_I|}{\sqrt{(\sigma_G^2 + \sigma_I^2)/2}}$, where $\mu_G, \mu_I, \sigma_G, \sigma_I$ are the

means and standard deviations of the genuine and imposter distributions, respectively [58][59]. *Daugman* reports that the decidability index in the case of *IrisCodes*[®] is as high as 11.36, which he claims to be “*much higher than that reported for any other biometric*” [58].

In [57][60] *Jain et. al.* offer a systematic treatment to the problem of individuality of fingerprints. Location and direction of fingerprint minutiae are used as independent features (independence in both senses that location and direction of a minutia are independent and

also these properties of different minutiae are independent). Minutiae are also assumed to only exhibit rigid-body deformations.

Jain *et. al.* [57][60] perceive individuality in practical terms as a problem of *similarity between intra-class patterns*, and hence they develop a model to study fingerprint individuality in this sense. Their model accommodates partial representations of a fingerprint and estimates the probability that given an overlap area (A) between two fingerprints, which permits existence of up to M minutiae “containers” partially filled by m minutiae from a temple fingerprint and n minutiae from an input fingerprint, the probability that at least q minutiae points are matched both in location and direction is given by:

$$p(q | M, m, n) = \sum_{\rho=q}^{\min(m,n)} \binom{m}{\rho} \binom{M-m}{n-\rho} \binom{\rho}{q} l^q (1-l)^{\rho-q} \bigg/ \binom{M}{n}.$$

Jain *et. al.* [57][60] estimate the parameters of their model based on empirical studies conducted on two large databases of fingerprints. To estimate M they first estimated a *match distance* threshold (r_o) by studying the distribution of distances between manually matched minutiae points from 450 matched pairs of fingerprints, and seeking the distance value below which most of the samples fall. They estimate M as $\left\lceil \frac{A/w}{2r_o} \right\rceil$, where w is the ridge period.

Similarly, they determine an *angular misalignment* threshold (θ_o) below which most matched minutiae points fall. They determined that for a fingerprint sensor resolution of 500 *dpi*, $r_o = 15$ pixels and $\theta_o = 22.5^\circ$. The value of θ_o then used to determine the parameter l . Finally the parameters m and n are determined statistically by averaging the distribution of the number of minutiae pints over the overlap area.

Jain *et. al.* [57][60] conduct a comparative study between accuracy estimates based on their individuality model and the performance metrics of state-of-the-art matchers on one hand, as well as between their individuality model and earlier fingerprint individuality models [61] on the other hand. They conclude that while the performance metrics of the state-of-the-art matchers are lower than predicted by their individuality model, their model still provides closer bounds than those provided by earlier models. They explain the looser bounds of the earlier studies as being a result of overlooking the intra-class variations, which lead to overestimating individuality (underestimating the probability of *false match*).

In [62] *Ratha et. al.* model the probability of *guessing* a fingerprint minutiae template as part of a study on enhancing security in biometrics-based authentication systems. Similar, to [57][60] they assume that a fingerprint template has discrete locations where minutiae may fit, however, they define a limited number of directions that may be assumed by a minutiae point, and they study the security strength of this representation. They present results suggesting that with 400 possible minutiae containers (20×20 template), given 40 reference minutiae, granting access to a system based on matching 25 subject minutiae is as strong (from a security perspective) as granting access based on a 16-character random alphanumeric password.

The problem of performance limits in biometrics has been stated and tackled using different approaches, and in the context of different applications. A common objective of all these studies is to predict performance limitations of biometric identification systems.

2.3 Problem Statement

We will study three problems inherently related to automation of the dental identification process:

- The first problem is on automatic classification of teeth into incisors, canines, premolars and molars as part of creating a data structure that guides tooth-to-tooth matching. We seek to classify teeth as such in order to avoid illogical comparisons between teeth, which inefficiently consume computational resources and may also mislead match decision-making.
- The second problem is on building a scheme for micro and macro decision-making, which exploits teeth multiplicity. Dental radiographic films not only capture projections of distinct teeth, but they may also provide multiple views for each of the distinct teeth. We want to utilize these aspects of teeth multiplicity in making robust record-level match decisions.
- In the third problem we study the performance limits of dental identification due to features capabilities. We consider two types of features used in dental identification, teeth contours and appearance features.

While the first problem was addressed in [14] and [15], both approaches use teeth contour features, automatic extraction of teeth contour is computationally intensive, and to the best of our knowledge, none of the techniques exploited in teeth contour extraction, so far, achieves this task reliably. None of the techniques presented in [11], [13], and [16] tackle the second problem as stated, i.e. none of these works surpass the level of film-to-film matching, nor do these address teeth view multiplicity in a wide sense. Finally, we know of no other attempts to study the limits on the performance of automated dental identification due to features capabilities.

In the following sections we formulate the three problems and where possible we use formal notations.

2.3.1 Problem One: Automatic Construction of Dental Charts

- **Given:**
 - A database of labeled dental Regions of Interest (ROIs) each assigned a class label $\omega_k \in \{ 'I', 'C', 'P', 'M' \}$, where '*I*' stands for Incisor, '*C*' stands for Canine, '*P*' stands for Premolar, and '*M*' stands for Molar.
 - A dental record D , with f radiographic films,
 - Each film of D contains Q dental ROIs,
 - Each of the Q dental ROIs corresponds to a tooth (t_q),
- **The Problem:**
 - To assign a tooth-class label ω_k to each t_q ,
 - To assign an intra dental quadrant number r_q ($r_q = 1, \dots, 8$) to each of the Q teeth of film f ,
 - To associate each dental ROI to a cell of the 32-cell, 4-quadrant dental chart, whose segments are $\{U_m, m = 1, \dots, 4\}$ and whose structure is shown in Figure 4.

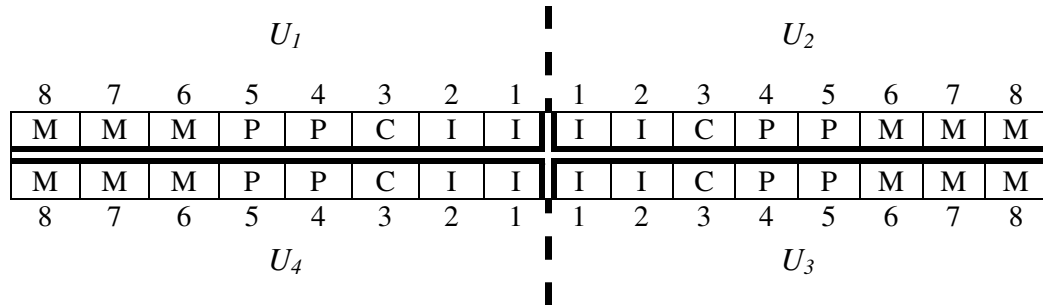


Figure 4: Structure of a 32-cell, 4-quadrant dental chart.

Note that there is a 1-to-many association between cells of the dental chart and dental ROIs, and that there is a 1-to-1 association between a dental ROI and a cell in the dental chart. Also note that this charting scheme does not follow the Universal Numbering System, which assigns a number between 1 and 32 to each tooth.

2.3.2 Problem Two: Micro and Macro Decision-Making

- **Given:**
 - A Subject dental record S with N_S teeth,
 - Each subject tooth (t_i) has $n(i)$ views,
 - A Reference dental record R with N_R teeth
 - Each reference tooth (τ_j) has $\eta(j)$ views,
 - Teeth t_i and τ_j are assigned identical intra dental quadrant number r_q ($r_q = 1, \dots, 8$),
 - Comparisons of the subject teeth set to the reference teeth set are to preserve inter and intra dental quadrant relations of both subject and reference dental charts (i.e. there are up to 4 possible configurations for inter dental quadrant comparisons between a subject/reference record-pair – out of which one configuration is legitimate. The legitimate configuration may not be known and each configuration involves comparison between up to 32 cells of the dental chart),
 - The probability of match (match score) $p_M(t_{ik}, \tau_{jl})$ between subject tooth t_i and reference tooth τ_j computed using the k^{th} view of t_i and the l^{th} view of τ_j , ($i = 1, \dots, N_S, j = 1, \dots, N_R, k = 1, \dots, n(i) l = 1, \dots, \eta(j)$).
- **The Problem:**
 - To determine the probability of match $p_M(t_i, \tau_j)$ between subject tooth (t_i) and reference tooth (τ_j) and hence make a decision $d_M(t_i, \tau_j)$ on the match status of t_i and τ_j ,
 - To use $\{d_M(t_i, \tau_j)\}$ as basis for making a decision $D_M(S, R)$ on the match status of S and R .
 - To compute a ranking score $\rho_M(S, R)$ that helps in sorting the resulting match list of the subject record S .

2.3.3 Problem Three: Performance Limits of Dental Identification

- **Given:**
 - A database of labeled dental Regions of Interest (*ROIs*) each assigned a class label $\omega_k \in \{ 'T', 'C', 'P', 'M' \}$, where '*T*' stands for Incisor, '*C*' stands for Canine, '*P*' stands for Premolar, and '*M*' stands for Molar,
 - Each *ROI* is associated with a unique *case ID* linked to the identity of the *ROI* bearer,
 - Two sets of dental features:
 - Teeth contour points, and
 - Teeth appearance features extracted using the scheme described in further detail in [6][21] and in the *Appendix* (see page 104),
- **The Problem:**
 - To estimate limits on the performance of dental identification due to each of the two named feature sets.

Chapter 3: Automatic Construction of Dental Charts

Automatic classification of teeth into incisors, canines, premolars and molars -and hence automatic construction of dental charts- plays a central role in guiding the process of tooth-to-tooth matching. Dental charts reflect the position of the radiographed teeth in the mouth, thus they help avoiding illogical teeth comparisons that inefficiently consume the limited computational resources and may also mislead the decision-making process.

In this chapter, we present an approach for automatic construction of dental charts that relies on Principal Component Analysis (PCA) and string matching techniques. Given a dental radiographic film annotated with teeth segmentation information [9][10][11][12], we first project the segmented teeth onto the subspaces of the four teeth classes, and then call the respective teeth classes that achieve the minimum classification risks. As a dental film usually shows multiple teeth, we exploit teeth neighborhood constraints to validate and correct the initially assigned teeth-classes. After validation and correction, we assign each tooth a number corresponding to its location in a dental chart.

3.1 Overview

Our approach for constructing dental charts relies on low computational-cost, appearance-based features to initially independently classify each segmented tooth of a dental film, i.e. without considering neighboring teeth. Thus, a tooth t_q is initially assigned a class label $\omega_k \in \{ 'I', 'C', 'P', 'M' \}$, where ' I ' stands for an Incisor, ' C ' stands for a Canine, ' P ' stands for a Premolar, and ' M ' stands for a Molar. Our approach is based on the assumption that a 2D view of a tooth crown is sufficient to determine the tooth class, which conforms to the practices followed by dentists.

To validate and possibly correct initial teeth class assignments, we use teeth class-neighborhood rules that govern the assortment of teeth in a human mouth. Finally, if the resulting teeth class sequence is unique (e.g. ' MP ', ' PC ', or ' CI ') we assign a number for each tooth corresponding to its position in the dental quadrant it belongs to. Otherwise, if the resulting teeth class sequence is non-unique (e.g. ' M ', ' MM ', ' MMM ', ' P ', ' PP ', ' I ', ' II ', or ' III ') we call a reject option.

Figure 5 shows the two stages involved in our approach for teeth classification. In the first stage we project a segmented tooth onto four image subspaces (or eigen-spaces) corresponding to the four teeth classes, then using a classification scheme we assign an initial class label for each segmented tooth. The second stage considers the neighborhood relations between the segmented teeth to validate and, if necessary, correct the initially assigned classes using teeth assortment rules as a prior and measurement clues from the first stage.

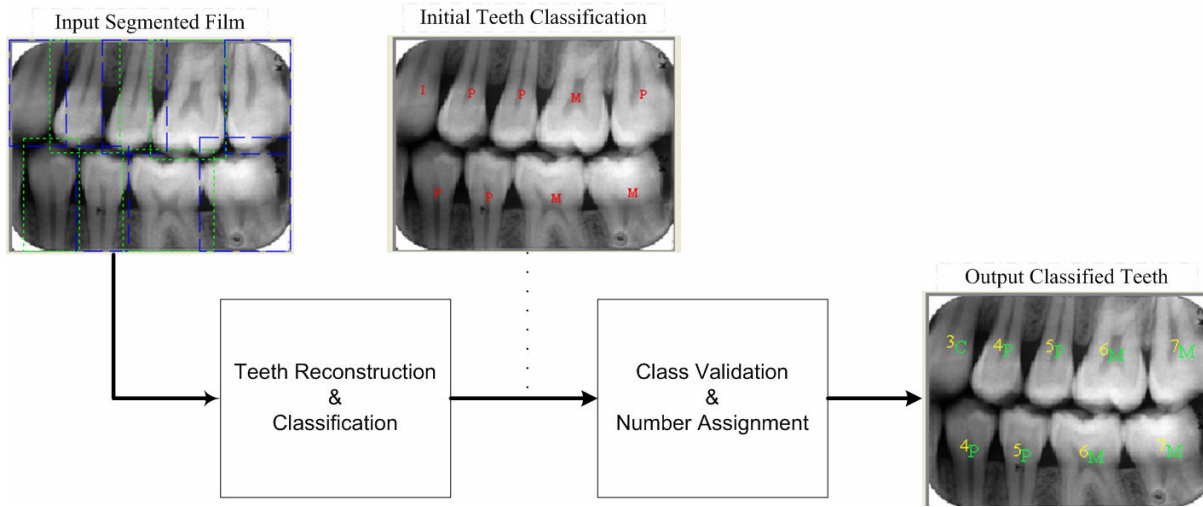


Figure 5: Block diagram of the two stages of teeth classification.

In section 3.2 we describe the first stage in fine detail and explain the underlying Principal Component Analysis (PCA) we use in constructing the four teeth subspaces. In section 3.3 we introduce the techniques of class validation and number assignment. In section 3.4 we present our experimental results and finally in section 3.5 we discuss the results and qualitatively compare our approach against two other approaches to the same problem.

3.2 Teeth Reconstruction and Classification

The *Ugly Duckling Theorem* suggests that “*there is no problem-independent or privileged or “best” set of features or feature attributes*” [41]. However, in designing a classification system, there are many factors that determine our choice of a set of features. The computational cost and reliability of automatic feature extraction are two key factors that determine our choice of practically feasible features.

By visually inspecting images of radiographed teeth of the four different classes, one can easily find similarities between teeth of the same class and discrepancies between teeth of

different classes. These similarities (and differences) are primarily broadly manifested in teeth peripheries. However, reliable automatic extraction of teeth contours is a very difficult problem and its current solutions are both inaccurate and computationally expensive. This motivates us to consider alternative appearance-based, low-cost features.

The key idea behind the initial stage of classification in our approach is to establish four image subspaces corresponding to the four teeth classes, and use the projections of a novel tooth onto these subspaces as basis for classification. To construct these subspaces we use exemplars from each tooth class as we discuss in section 3.2.1. With these image subspaces constructed, we achieve initial teeth classification as depicted in Figure 6:

- An input tooth t_q is view-normalized, as we describe in section 3.2.2, to compensate for possible geometric and photometric variations that may cause significant differences between that tooth and the exemplar sets used for constructing the four subspaces.
- The view-normalized input tooth t_{qN} is projected onto the four image subspaces, as we describe in section 3.2.3. Hence, we obtain four coefficient sets $\{w_I\}$, $\{w_C\}$, $\{w_P\}$, and $\{w_M\}$, corresponding respectively to the projections of t_{qN} onto the *Incisors* subspace, the *Canines* subspace, the *Premolars* subspace, and the *Molars* subspace.
- The obtained weight sets are used in conjunction with the sample mean of each of the four teeth classes to reconstruct the view-normalized tooth t_{qN} in the four image subspaces, thus obtaining the approximations $\hat{T}_I, \hat{T}_C, \hat{T}_P, \hat{T}_M$.
- We feed t_{qN} and each of its four approximations to classifier that calls out one of the four classes according to least energy discrepancy between the view normalized tooth and its four approximations, thus obtaining an initial class assignment for t_{qN} as we describe in section 3.2.3.

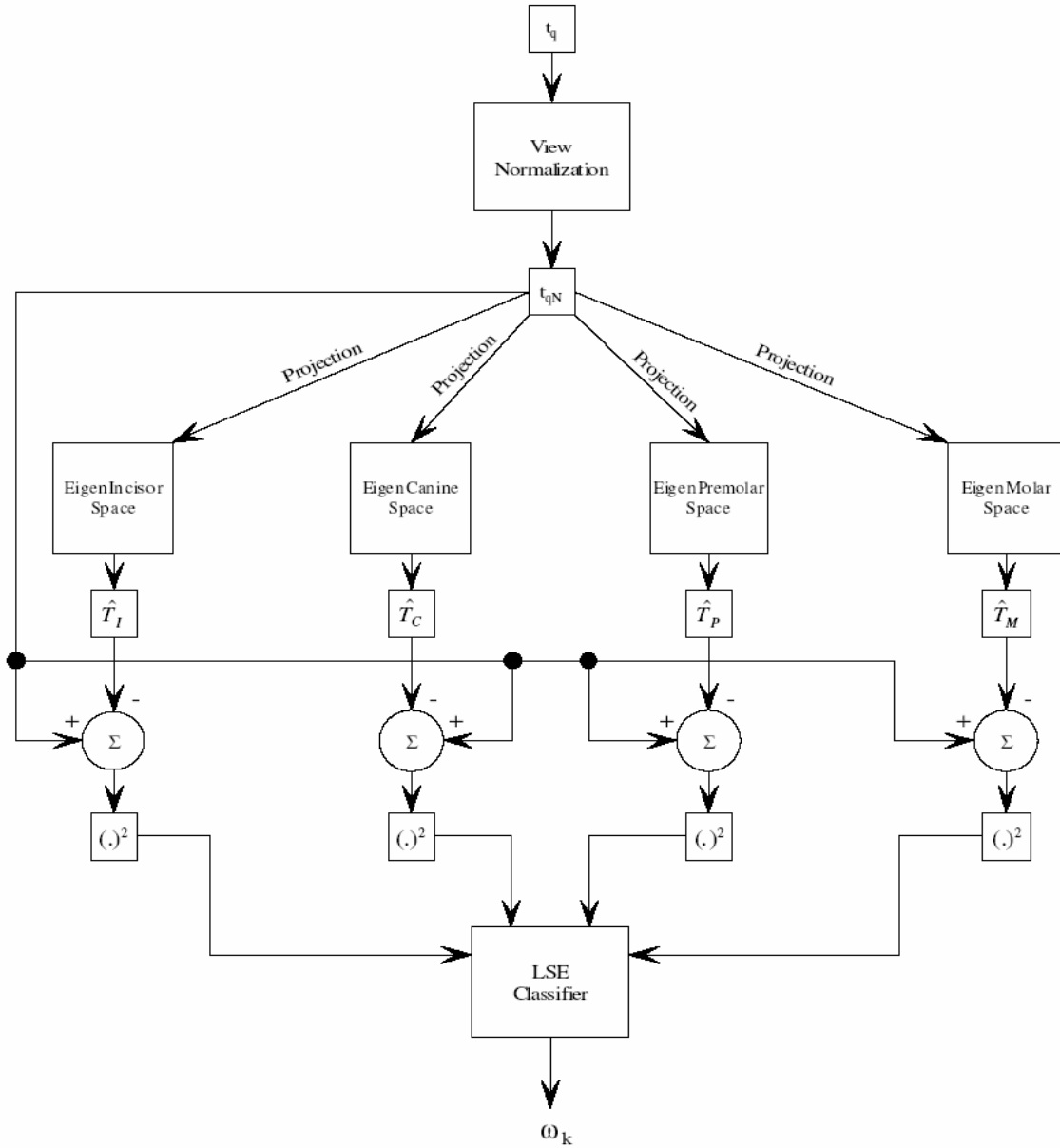


Figure 6: Teeth reconstruction and initial classification.

3.2.1 The Eigenteeth

The heart of our approach for initial teeth classification is the assumption that we can establish four image subspaces, each of which captures the most common appearance features shared between typical instances of each of the respective teeth classes. To construct these image subclasses we resort to Principal Component Analysis (PCA). Because of the some resemblance between our approach for teeth reconstruction and the *eigenfaces* approach [46] (at least in the abstraction level), we refer to the resulting basis vectors

defining the image subspaces as the *eigenteeth*. We first select representative teeth samples of each of the teeth classes. Figure 7 depicts a sample of exemplars from each teeth class.

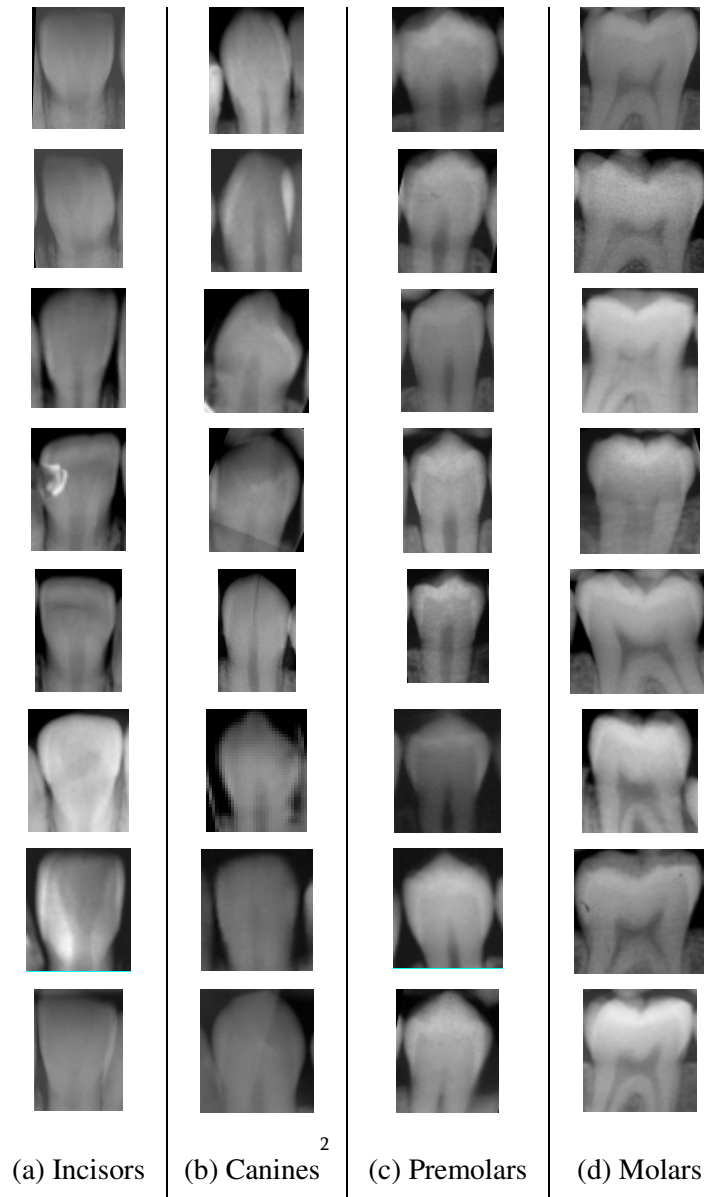


Figure 7: Sample of teeth used in constructing the image subspaces of the four teeth classes.

The next step towards constructing the image subspaces is to confine the selected samples to a square area that bounds teeth lateral surfaces such that teeth crowns are slightly below the

²

Dr. Robert Howell noticed that four of the canine samples shown in this figure are actually premolars, *see* section 3.5 for a note on the apparent similarity between instances of the different teeth classes.

topside of the square area. This confinement step is aimed to reduce the view variations between exemplars of the same class. The resulting square images are then resized to a fixed sized image of 32x32 pixels using the *bicubic* interpolation scheme.

We augment the exemplar sets with conjugates of the exemplars produced by horizontal flipping (as depicted in Figure 8), i.e. around a vertical axis. The value of this is to substitute for orientation biases in the sample, thus in effect we end up with equal number of exemplars from the left and side quadrants. Moreover, to reduce the complexity of the problem, all maxillary plane exemplars are vertically flipped such that their crowns appear on the top of the image.

With the sets of exemplars prepared, we proceed with the construction of the subspaces as follows:

- Compute the average of the four samples: $\underline{\mu}_{\omega_k} = \frac{1}{N_{\omega_k}} \sum_{i=1}^{N_{\omega_k}} \underline{x}_{i\omega_k}$, $\omega_k \in \{'I', 'C', 'P', 'M'\}$ where N_{ω_k} is the number of exemplars of class ω_k , and $\underline{x}_{i\omega_k}$ is the vector representation of the i^{th} ω_k exemplar,
- Let $A_{\omega_k} = [(\underline{x}_{1\omega_k} - \underline{\mu}_{\omega_k}) \dots (\underline{x}_{i\omega_k} - \underline{\mu}_{\omega_k}) \dots (\underline{x}_{N_{\omega_k}} - \underline{\mu}_{\omega_k})]$ denote the matrix of the average-adjusted sample of class ω_k , then we compute the eigenvectors $\underline{u}_{i\omega_k}$ of the scatter matrices of the four samples using the efficient technique proposed in [46] by first solving the eigenvalue problem $A_{\omega_k}^T A_{\omega_k} \underline{v}_{i\omega_k} = \lambda_{i\omega_k} \underline{v}_{i\omega_k}$, where $\lambda_{i\omega_k}$ is the i^{th} eigenvalue of $A_{\omega_k}^T A_{\omega_k}$. Finally, we linearly map $\underline{v}_{i\omega_k}$ to $\underline{u}_{i\omega_k}$ using $\underline{u}_{i\omega_k} = A_{\omega_k} \underline{v}_{i\omega_k}$.
- Choose K eigenvectors for each set whose corresponding eigenvalues are the largest.

Figure 9 shows image representations of the first 15 eigenvectors of each of the four subspaces resulting from the PCA of four sample sets. The image (in green) on the upper left corner of each set represents the sample average $\underline{\mu}_{\omega_k}$, $\omega_k \in \{'I', 'C', 'P', 'M'\}$ of the respective set. The eigenvectors arranged according to their corresponding eigenvalues such that the first *eigenincisor*, for instance, is to the immediate right of the average incisor and so on.

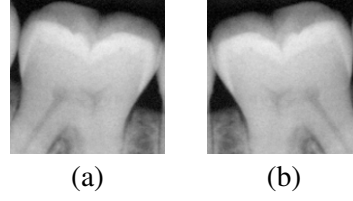


Figure 8: (a) Upward pointing crown (b) Conjugate of (a) under horizontal flipping.

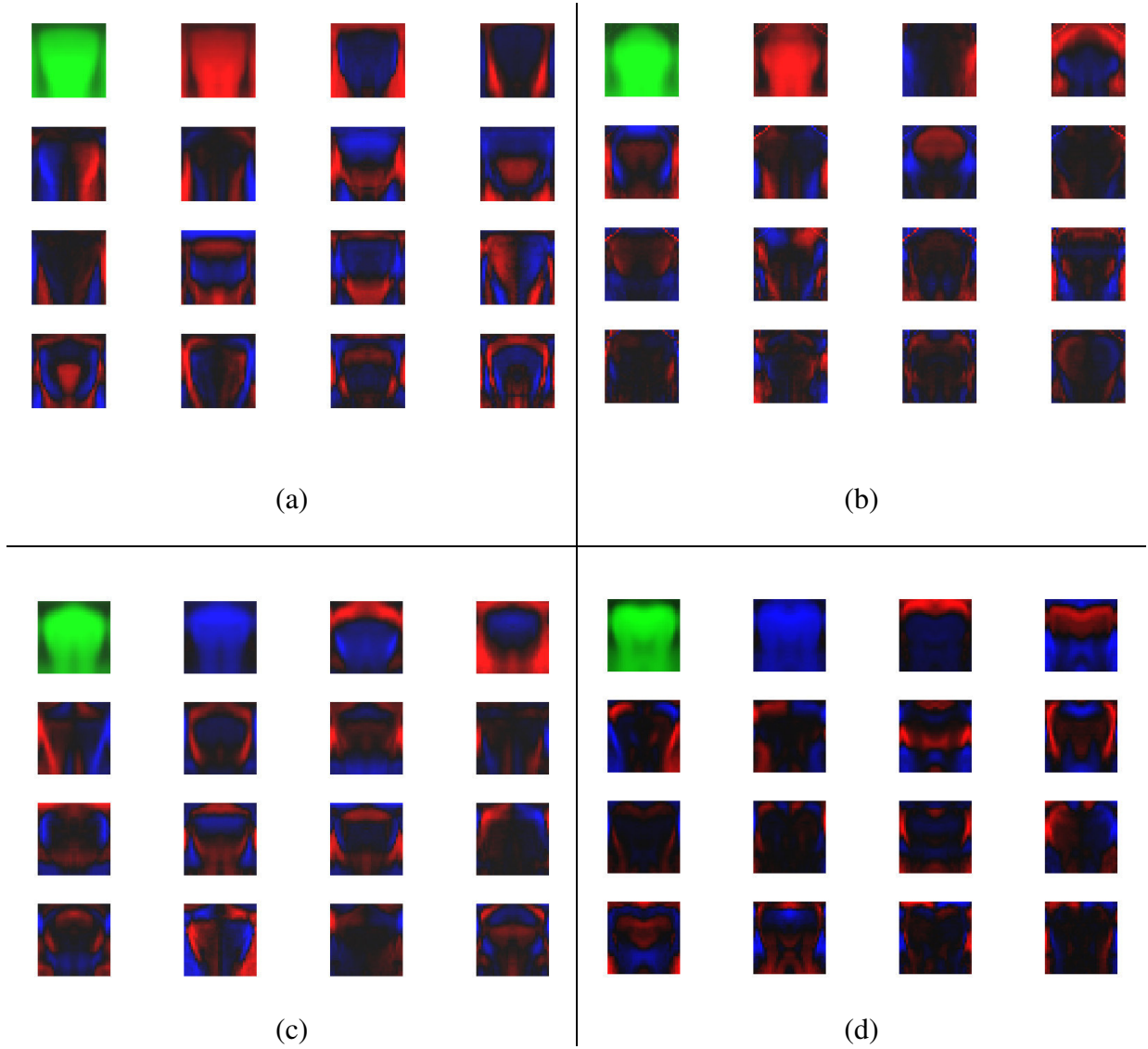


Figure 9: Image representation of (a) *EigenIncisors*, (b) *EigenCanines*, (c) *EigenPremolars*, and (d) *EigenMolars*.

3.2.2 View Normalization

In practice, dental radiographic films show teeth that need not comply with the standardizing assumptions on view that we made. On one hand, crowns need not appear on the topside of

the segmented regions, as in the case with upright films showing teeth of the upper jaw or flipped films showing teeth of the lower jaw. On the other hand, the lateral surfaces of radiographed teeth do not necessarily appear as predominantly vertical lines.

While some of the teeth segmentation algorithms annotate teeth as either upper or lower, and thus partly solve the first type of deviations from standard view, other teeth segmentation algorithms do not discriminate between upper and lower teeth. To accommodate the different teeth segmentation algorithms, we conduct a simple test that advises on whether the crown is closer to the top or the bottom of the segment, and accordingly we may flip the segment vertically. Figure 10 (a) depicts gradient vectors of the test pattern that we use to decide whether a segment should be flipped. The rationale for choosing this pattern is that: a segmented tooth, whose crown is mostly oriented upwards like that shown in Figure 10 (b), would have a larger sum of the dot product of its intensity gradient vectors with the gradient vectors of the upper zone of the test patterns than the sum of the dot product due to the lower zone. And likewise a segmented tooth, whose crown is mostly oriented downwards like that shown in Figure 10 (c), would have a larger sum of the dot product due to the lower zone than that due to the upper zone.

In view normalization, there are two types of transformations that we usually want to address; the first is the geometric transformations, which we treat in the remainder of this section. The second type of transformations is the photometric (or intensity) transformation, which we address at the earlier stage of film enhancement as part of record preprocessing [9].

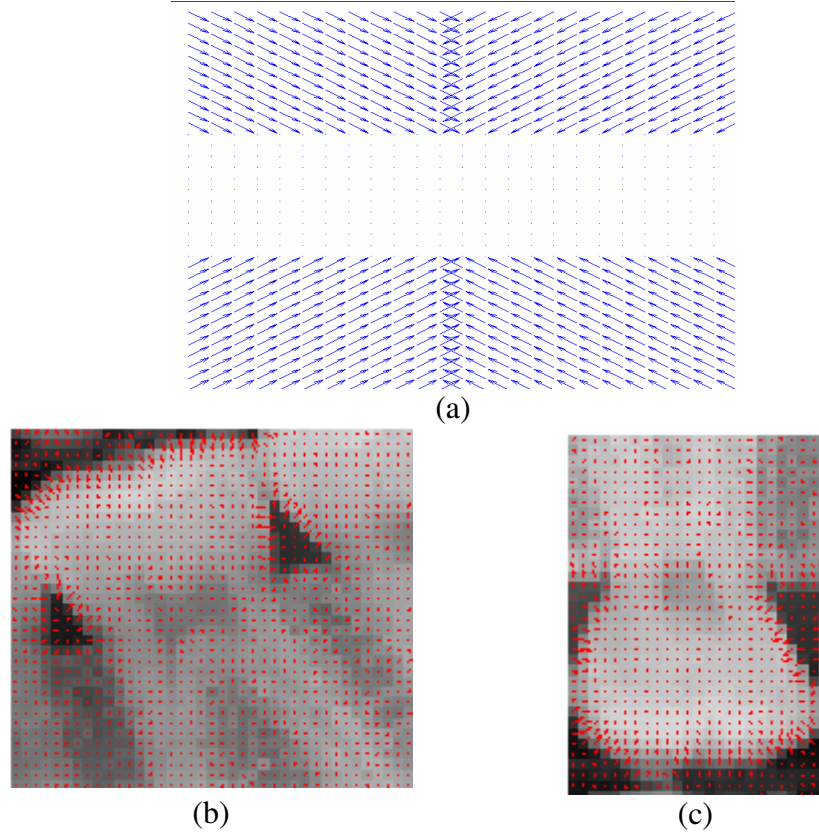


Figure 10: The gradient vectors of (a) the test pattern, (b) a segmented tooth with upward crown, and (c) a segmented tooth with downward crown.

The next step towards normalizing the view of a segmented tooth is to ensure that its lateral surfaces appear predominantly vertical. Starting with a tooth with an upward oriented crown as an input image, e.g. as in Figure 11 (a), we complete the view normalization task in the following steps:

- We obtain an edge image (as in Figure 11 (b)) from the input image using *Canny* edge detection [63]. Our choice of the width of the smoothing kernel is such that to avoid any excessive number of edge points due to intensity variations not related to the tooth surfaces. We further group edge points based on connectivity and remove short edge segments.
- Using the gradients of the input image, we estimate the orientation at each of the remaining edge points, and accordingly classify the orientations of the edge points as either horizontal or vertical. Hence we compute a histogram of orientations of the vertical edge points (as in Figure 11 (c)).

- We compute the mean of the vertical orientations, and use its negative as the angle by which we rotate the input image such that the lateral surfaces of the tooth appear predominantly vertical (as in Figure 11 (d)).
- Using a test pattern similar to that shown in of Figure 10 (a) but with a non vacant middle zone, we retain the line segments of the rotated edge image that have inward intensity gradient (as in Figure 11 (e)). These line segments are more relevant to the tooth periphery and thus help in confinement of the tooth ROI.
- By analyzing the vertical and horizontal span of the tooth, we confine our tooth ROI to a square area (as in Figure 11 (f)), whose top is just above the predominantly horizontal line segment belonging to the crown and such that the lateral sides of the tooth are also contained in the square area.
- The last step is to resize the tooth ROI to a 32×32 image, and thus we get t_{qN} (as in Figure 12), the view normalized version of the segmented tooth t_q .

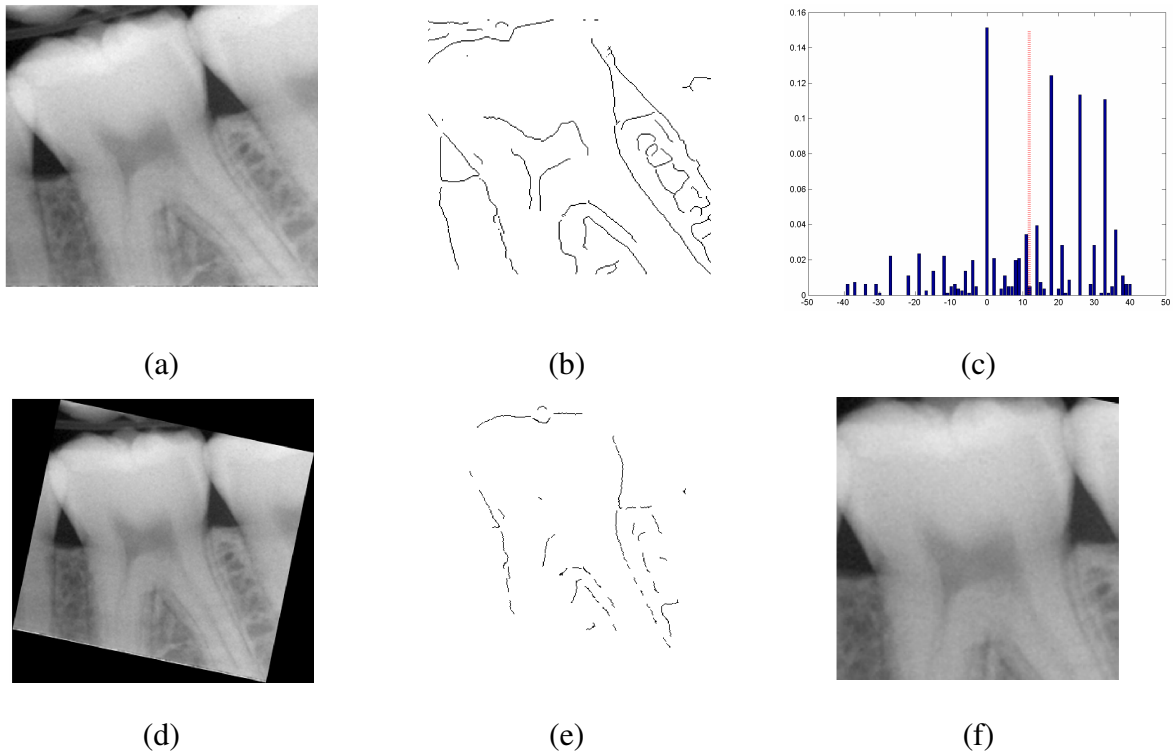


Figure 11: Illustration of the orientation adjustment and ROI confinement steps of a segmented tooth.

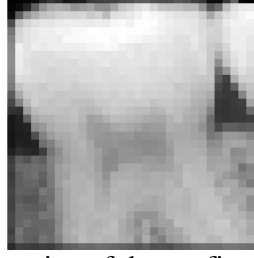


Figure 12: Image representation of the confined tooth ROI after resizing.

3.2.3 Projection and Classification

Now that we obtained a view normalized tooth t_{qN} , we project it onto each of the four teeth-class subspaces to obtain four sets of coefficient, namely $\{w_I\}$, $\{w_C\}$, $\{w_P\}$, and $\{w_M\}$, which respectively represent the projections of t_{qN} on the Incisors subspace, the Canines subspace, the Premolars subspace, and the Molars subspace. The coefficient sets are given by: $\{w_{\omega_k}\} = (t_{qN} - \underline{\mu}_{\omega_k})^T \times [\underline{u}_{1\omega_k} \ \underline{u}_{2\omega_k} \ \dots \ \underline{u}_{i\omega_k} \ \dots \ \underline{u}_{K\omega_k}]$, $\omega_k \in \{'I', 'C', 'P', 'M'\}$, where $\underline{\mu}_{\omega_k}$ is the mean vector of class ω_k , and $\underline{u}_{i\omega_k}$ is the i^{th} eigenvector of that class.

Using the respective coefficients and the training sample mean images the four teeth classes, we reconstruct t_{qN} in each of the subspaces, thus obtaining its four approximations according to: $\hat{T}_{\omega_k} = \underline{\mu}_{\omega_k} + [\underline{u}_{1\omega_k} \ \underline{u}_{2\omega_k} \ \dots \ \underline{u}_{i\omega_k} \ \dots \ \underline{u}_{K\omega_k}] \times w_{\omega_k}^T$, $\omega_k \in \{'I', 'C', 'P', 'M'\}$. Figure 13 shows an example of tooth reconstruction in the four subspaces, where we approximate the view-normalized molar shown in Figure 12 to its images on each of the subspaces. Notice that among the four images (Figure 13 a-d), the closest in appearance to the image of this molar is its approximation using the Molars subspace. It is worth mentioning that the molar image shown in Figure 12 was not used in constructing the Molars subspace.

The next step is to determine the energy discrepancy between t_{qN} and its approximations, which we determine as follows: $\epsilon_{\omega_k} = \sum (\hat{T}_{\omega_k} - t_{qN})^2$, $\omega_k \in \{'I', 'C', 'P', 'M'\}$, where (x, y) are image coordinates. Figure 14 shows image representations of the difference between the view-normalized molar shown in Figure 12 and its teeth class approximations shown in Figure 13. Notice that the approximation of this molar resulting from its reconstruction in the Molars subspace produced darker spots than its approximations to the other teeth classes.

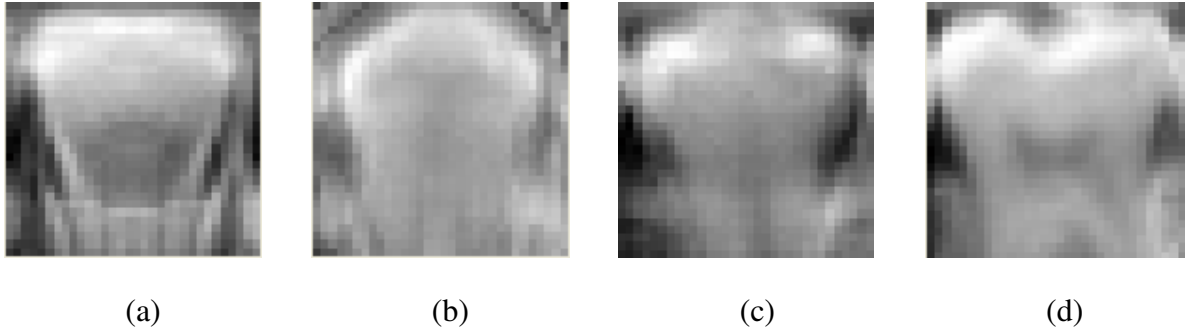


Figure 13: Reconstructed image of the view normalized molar (Figure 12) in the four subspaces.

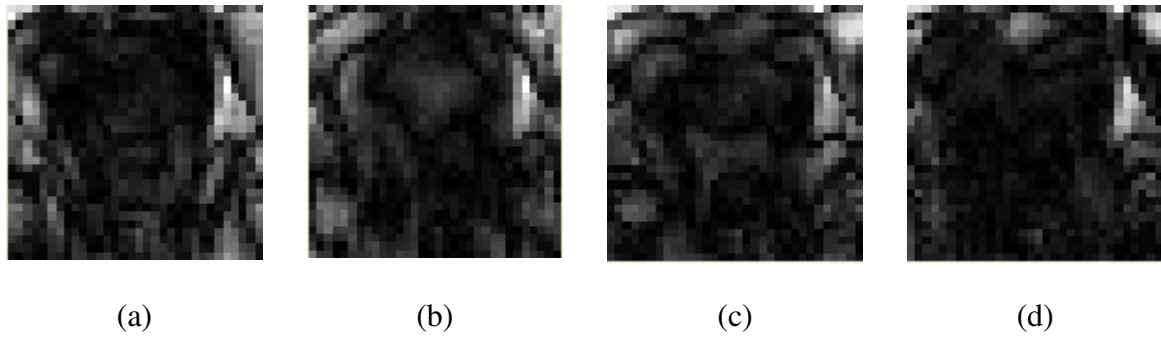


Figure 14: Scaled error images between the molar (Figure 12) and its images in the four subspaces.

We assign tooth t_q to class ω_k based on the least energy discrepancy rule as follows:

$$t_q \rightarrow \omega_k \mid k = \arg \min_{<j>} \{\mathcal{E}_j\}.$$

3.3 Class Validation and Number Assignment

As in most of the classification problems, the initial class labels assigned to each tooth, according to the least energy discrepancy rule, are prone to errors. However, a dental film usually shows a number of teeth, and because the assortment of teeth in a human mouth follows a specific pattern, we rely on teeth neighborhood rules to validate the detected sequence of teeth class labels. We also seek to correct, if possible, sequences that do not conform to the reference pattern of possible sequences. And finally if we find the validated/corrected sequence unique, we assign a number to each tooth corresponding to its position in its dental quadrant.

Our approach for class validation is based on string matching. When validating bitewing sequences, we also take into consideration the horizontal distance between teeth in the upper and lower jaws. So if we let:

- \mathbf{X} denote the 16 character reference string ‘MMMPPCIIIICPPMMM’,
- $\mathbf{S_F} = \{s_1 s_2 \dots s_j \dots s_n, 1 \leq n \leq 16 \mid s_j \in \{‘I’, ‘C’, ‘P’, ‘M’\}\}$ denote the sequence of the initially assigned labels of the segmented teeth of the radiographic film F ,

We post the class validation problem as a string-matching problem with error [41], where we seek to match the pattern $\mathbf{S_F}$ to the text \mathbf{X} with the possibility of error in the former. Because there may be different ways to correct an error, we compute the cost for each error correction possibility.

In section 3.2.3 we used $\{\epsilon_k, k=1, \dots, 4\}$ to denote the energy discrepancies between the image of a view-normalized tooth and its four approximations. We define the cost $c_{k_o \rightarrow k}^{[i]}$ of changing the class label of the i^{th} tooth from its initially assigned class k_o to the label of class

k as follows: $c_{k_o \rightarrow k}^{[i]} = \frac{\epsilon_k^{[i]} - \epsilon_{k_o}^{[i]}}{\max_{<j>}(\epsilon_j^{[i]}) - \epsilon_{k_o}^{[i]}}$. And accordingly we define the total cost $C_{S_F \rightarrow S_F}$ for

changing the pattern $\mathbf{S_F}$ to $\hat{\mathbf{S_F}}$ as $C_{S_F \rightarrow \hat{S_F}} = \sum_{<i>} c_{k_o(i) \rightarrow k(i)}^{[i]}$. Of all the possible changes, if a change is required due to impossibility of matching $\mathbf{S_F}$ to \mathbf{X} without errors, we seek $\hat{\mathbf{S_F}}$ to minimize the cost (or *edit distance* [41]) $C_{S_F \rightarrow \hat{S_F}}$.

There are instances where a match exists between $\hat{\mathbf{S_F}}$ (or $\mathbf{S_F}$) and \mathbf{X} at more than a unique shift, e.g. if $\mathbf{S_F}$ or $\hat{\mathbf{S_F}} \in \{‘MM’, ‘MMM’, ‘PP’\}$. This situation specifically appears when $\mathbf{S_F}$ (or $\hat{\mathbf{S_F}}$) exhibits no transitions between teeth classes. When we encounter such a situation with a periapical film, we declare a teeth-numbering tie. However, when we encounter such a situation with bitewing films, and if the sequence resulting from the opposite quadrant is unique, we attempt to determine a unique sequence. Moreover, with bitewing views we detect, and either correct or identify, instances where the resulting sequences of the upper and lower quadrants are inconsistent with one another, i.e. crisscrossed quadrants.

3.4 Experimental Results

To setup the image subspaces for the four teeth classes we prepared a dataset of exemplars using the dental image database provided by the Missing and Unidentified Person Unit of

Washington State Patrol [64]. We segmented and preprocessed 100 Molars, 100 Premolars, 30 Canines, and 60 Incisors from records.

In selecting these teeth we tried the best we could avoiding biased datasets. So we evenly used teeth from upper and lower jaws, we used teeth from the right and left sides of the mouth, we used teeth images of different intensity contrast. However, only 6.5% of the used teeth showed fillings (*see* Figure 7 for a sample of images used of constructing the image subspaces). We computed the eigenvectors for each of the covariance matrices due to the exemplar sets of the four classes.

To test our proposed approach for teeth classification and numbering, we prepared a test dataset disjoint with the dataset of exemplars described earlier. The test dataset was prepared using the CJIS ADIS project dental image database [65] and the test set prepared by Dr. Robert Howell [66]. Table 1 lists the number of films and teeth used in our testing experiments categorized by film type. Figure 15 depicts the overall distribution of teeth used in our testing experiments; ‘R’ and ‘L’ respectively denote the right and left face sides, ‘X’ and ‘D’ respectively denote the upper and lower jaws, and the numbers denote the intra-quadrant position of a tooth (*see* Figure 4). We notice that premolars and molars dominate the figure, because these teeth appear in dental films more often.

Table 1: The number of films and teeth used in testing of teeth classification.

	Number of films	Number of teeth
Bitewing films	269	1406
Lower Periapical films	134	358
Upper Periapical films	104	263

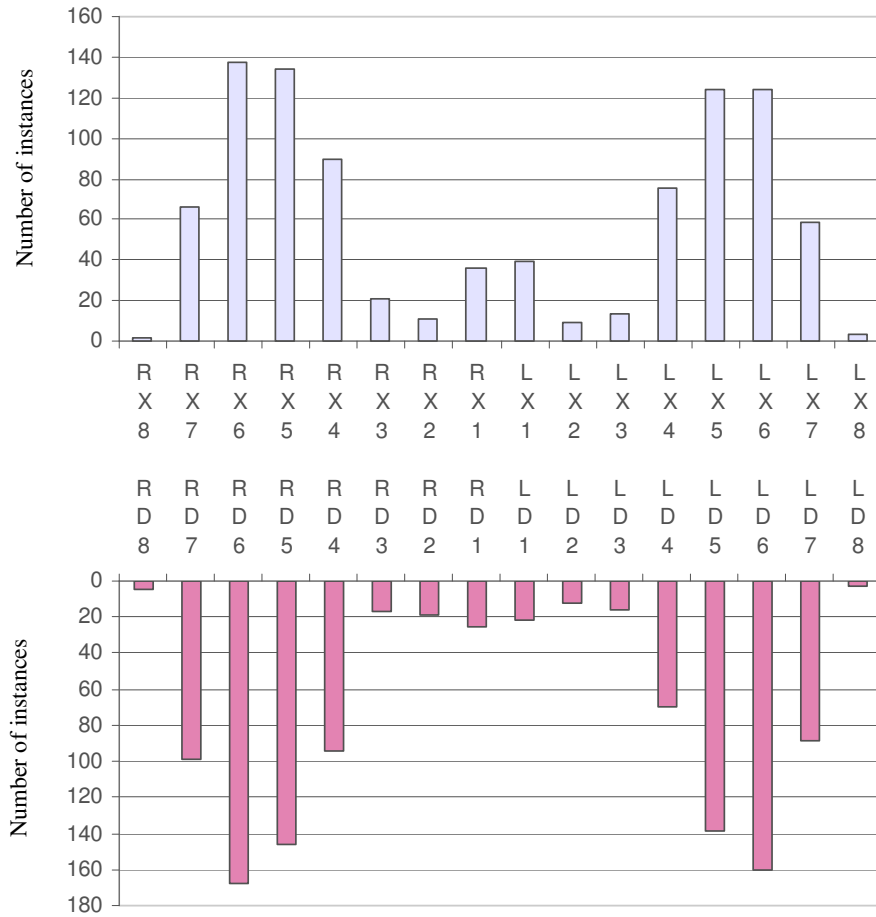


Figure 15: Teeth distribution of the test dataset used in teeth classification experiments.

We varied the number of eigenvectors used as basis for the four subspaces and found that keeping the first 15 eigenvectors per subspace achieves better performance compared to keeping 5, 10, 20, 25, or 40 eigenvectors. The more the number eigenvectors we use to define the subspaces, the higher the classification time. The results reported herein are thus based on experiments using 15 eigenvectors per subspace.

As we described in section 3.3, in some instances our class validation scheme declares a tie. The number of bitewing teeth declared in tie was 85 teeth (~ 6%); the number of lower

periapical teeth declared in tie was 148 teeth (~ 41%); and the number of lower periapical teeth declared in tie was 119 teeth (~ 45%).

The teeth classification confusion matrices for the three types of films we tested are given in Table 2 (all films), Table 3 (bitewing films only), Table 4 (lower periapical films only), and Table 5 (upper periapical films only). We observe the following:

- Overall, initial classification produced ~ 75% accuracy, teeth class validation raised the overall teeth classification accuracy to 86%. The accuracy of intra-quadrant numbering is ~ 70%, however, ~ 85% of assigned intra-quadrant numbers are within one position from the true value.
- In bitewing films, initial classification produced ~ 79% accuracy, teeth class validation raised the overall teeth classification accuracy to 90%. The accuracy of intra-quadrant numbering is ~ 76%, however, ~ 91% of assigned intra-quadrant numbers are within one position from the true value.
- In lower periapical films, initial classification produced ~ 59% accuracy, teeth class validation raised the overall teeth classification accuracy to 72%. The accuracy of intra-quadrant numbering is ~ 50%, however, ~ 64% of assigned intra-quadrant numbers are within one position from the true value.
- In upper periapical films, initial classification produced ~ 64% accuracy, teeth class validation raised the overall teeth classification accuracy to 76%. The accuracy of intra-quadrant numbering is ~ 46%, however, ~ 67% of assigned intra-quadrant numbers are within one position from the true value.

Table 2: Confusion matrix for teeth classification in all films

		Assigned Class			
		I	C	P	M
True Class	I	12	4	9	0
	C	4	2	21	3
	P	28	8	751	19
	M	5	2	311	496

(a) Initial Classification

		Assigned Class			
		I	C	P	M
True Class	I	11	5	8	1
	C	3	9	16	1
	P	5	15	752	34
	M	4	3	128	679

(b) After Validation

Table 3: Confusion matrix for teeth classification in bitewing films only

		Assigned Class			
		I	C	P	M
True Class	I	0	0	0	0
	C	0	0	12	0
	P	0	1	648	14
	M	0	0	249	397

(a) Initial Classification

		Assigned Class			
		I	C	P	M
True Class	I	0	0	0	0
	C	0	4	8	0
	P	0	3	638	22
	M	0	0	98	548

(b) After Validation

Table 4: Confusion matrix for teeth classification in lower periapical films only

		Assigned Class			
		I	C	P	M
True Class	I	11	0	5	0
	C	4	2	1	1
	P	26	6	47	2
	M	5	2	34	64

(a) Initial Classification

		Assigned Class			
		I	C	P	M
True Class	I	10	2	4	0
	C	3	4	0	1
	P	5	12	63	1
	M	4	3	23	75

(b) After Validation

Table 5: Confusion matrix for teeth classification in upper periapical films only

		Assigned Class			
		I	C	P	M
True Class	I	1	4	4	0
	C	0	0	8	2
	P	2	1	56	3
	M	0	0	28	35

(a) Initial Classification

		Assigned Class			
		I	C	P	M
True Class	I	1	3	4	1
	C	0	1	8	0
	P	0	0	51	11
	M	0	0	7	56

(b) After Validation

In Figure 16 we show examples of successful teeth numbering of bitewing films as well as examples of incorrect teeth numbering. In Figure 17 and Figure 18 respectively we show examples of correctly and incorrectly numbered in lower and upper periapical films. In Figure 19 we show an example of a rejected bitewing film due to a quadrant crisscross situation that we could not resolve. One other cause of ties is the non-uniqueness of a teeth sequence as we described in section 3.3.

We carried out these tests using un-compiled MATLAB[®] realizations running on a 256 MB RAM, 1.13 GHz Intel P3[®] based personal computer. The average tooth classification time was 0.3 seconds, and the average film classification time was 0.8 seconds.

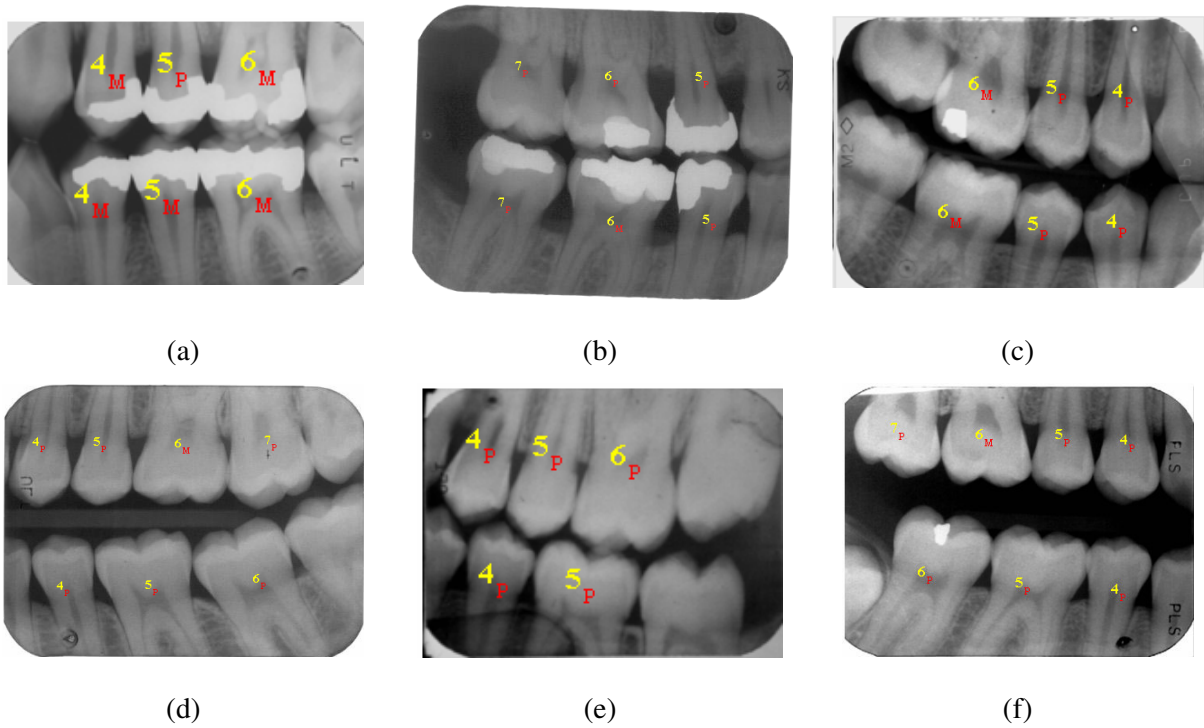


Figure 16: Examples of bitewing films (a – c) with correct teeth numbers (d – f) with incorrect teeth numbers.

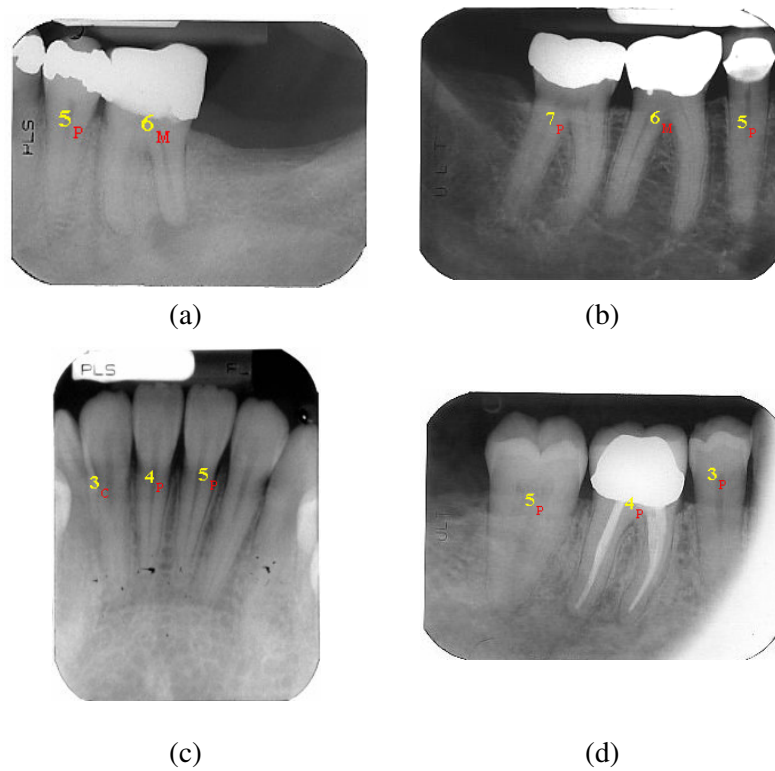


Figure 17: Examples of lower periapical films (a, b) with correct teeth numbers (c, d) with incorrect teeth numbers.

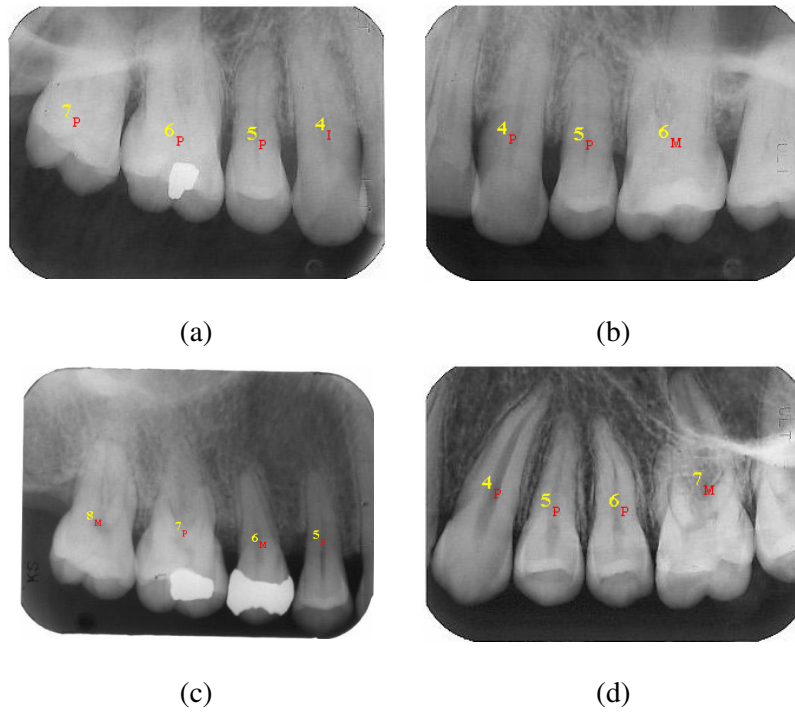


Figure 18: Examples of upper periapical films (a, b) with correct teeth numbers (c, d) with incorrect teeth numbers.

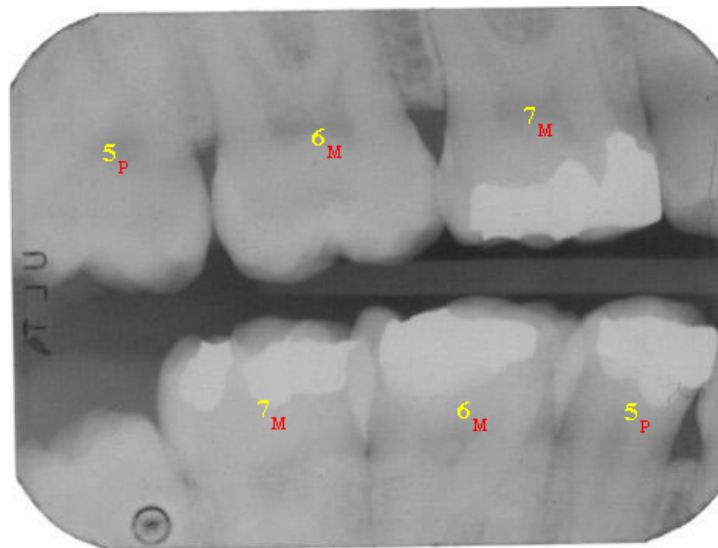


Figure 19: Example of a rejected bitewing film due to inverted upper and lower quadrants.

3.5 Discussion

We observe that Premolars and Molars are more often used for identification purposes. A careful look at the confusion matrix presented in Table 2 (a) reveals that, overall Premolars are initially highly correctly classified $\sim 93\%$, we also observe that misclassified Molars are predominantly seen as Premolars.

The performance of the class validation scheme, and hence the overall performance of teeth numbering and classification, depends on how well we assign the cost (or risk) of misclassification. We notice that accurate teeth numbering calls for proper detection of transitions between teeth classes, so for example if after class validation we achieve 97% accuracy in the classification of Premolars and 83% accuracy in the classification of Molars, then we are likely not to exceed 77% in the accuracy of teeth numbering.

In assigning the misclassification risks, we used the energy discrepancies between the view normalized images and their approximations. We could achieve an overall teeth-numbering accuracy of $\sim 76\%$ (an improvement of 5%) in the teeth numbering accuracy by adjusting this assignment, such that we take into consideration the observations we just made on the Molars misclassification rate, and the accuracy of Premolars classification. The teeth-numbering accuracy in bitewing films after adjusting the cost assignment scheme became $\sim 82\%$ (an improvement of 6%).

The adjustment we made is such that it always costs less to reassign a Premolar as a Molar, and is based on some educated guesses, which led to $\sim 97\%$ accuracy in the classification of Premolars and $\sim 89\%$ accuracy in Molars classification. We think, however, that more improvements can be made on the teeth-numbering accuracy by means of learning the misclassification risks.

We observed that the main sources for errors in initial classification are partly due to poor image contrast, e.g. Figure 20 (a), projection overlaps between neighboring teeth leading to imperfections in segmentation, e.g. Figure 20 (b), and inherent similarities in some cases Premolars and Molars, e.g. Figure 20 (c). While we did not notice severe degradation due to

fillings, we think that filling detection and substitution using image-inpainting³ techniques may also improve the performance of initial teeth classification. In many cases, there are also inherent appearance similarities between Canines and Premolars, that is why the classification scheme proposed in [15], *Jain* and *Chen* do not make class distinction between Premolars and Canines and they collectively classify them as (bi)cuspsids. One of the ideas to improve the accuracy of classification of premolars and molars is to take into consideration the pulp size, which is smaller in premolars as compared to molars.

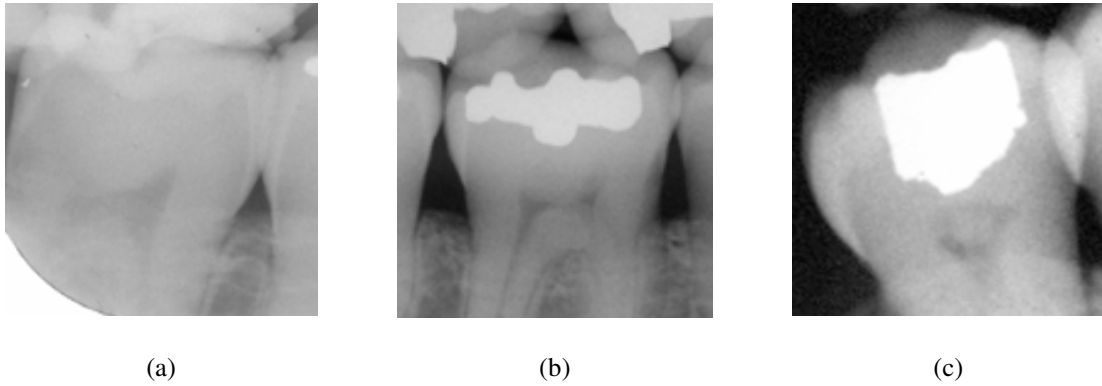


Figure 20: Examples of initially misclassified teeth.

In [14] *Mahoor* and *Abdel-Mottaleb* presented an approach for teeth classification in dental bitewing views using *Fourier* descriptors of teeth contours. The approach that they used for validation of sequences, does not fully exploit neighborhood relations in bitewing films, particularly they do not use inter-quadrant sequences for sequence correction and hence the numbers assigned to the different teeth on the upper/lower quadrants may include inconsistencies (e.g. quadrant crisscrossing). In [15], *Jain* and *Chen* also do not indicate that they use any inter-quadrant validation, which we think is important particularly in bitewing views.

The main advantage of the work presented by *Jain* and *Chen* in [15] is that by using a Hidden Markov Model for registration of teeth to a human dental atlas, they have implicitly embedded the capability of learning the proper misclassification costs that we pointed to

³ We attribute the observation on inpainting of teeth with fillings to Dr. Xin Li.

earlier. Another advantage of their work is that it also addresses the problem of teeth classification with missing teeth. We believe that with some extra segmentation post processing and slight modification of the string matching technique we use such that it also accommodates with “*Don’t Care*” symbols [41], our approach would be capable of handling classification with missing teeth. Some of the features to explore in determining whether a film indicates a missing tooth, if any, and the number of teeth missing are the size of the gap between teeth, as well as the tilt angles of teeth surrounding a gap, which would also depend on the type of teeth before and after the gap.

The main advantage of our technique over the works introduced in [14] and [15] is that we use inexpensive features with slight processing overhead, and achieve comparable results to those reported in [14] and [15], also keep in mind that we used a larger test dataset than those used in these works. To the best of our knowledge, reliable automatic extraction of teeth contours is a difficult problem that remains as an unsolved challenge. We think that practically, we may resort to these more expensive techniques in situations, where a tie is encountered given that the other two techniques may resolve the tie.

The training and testing experiments we performed are based on disjoint dental image datasets; and while both datasets are relatively large in size, it would be interesting to perform a cross validation study, where we use multiple subsets for training and others for testing. The outcome of these experiments would be confusion matrices along with confidence intervals, as opposed to frequencies only.

So far we assumed that the teeth classes used by dentists are the useful classes to exploit in automating the dental identification process. However, the most suitable classes from a machine perspective need not be the same as those used by humans. It would be interesting to consider machine-oriented teeth classes in future research.

3.6 Summary

Automatic classification of teeth into incisors, canines, premolars and molars, and hence automatic construction of dental charts, is essential for guiding the process of tooth-to-tooth matching. Dental charts capture the position of the radiographed teeth in the human mouth,

thus they help avoiding illogical teeth comparisons that inefficiently consume the limited computational resources and may also mislead the decision-making process.

In this chapter, we presented a dual-stage approach for automatic construction of dental charts based on low computational-cost, appearance-based features and string matching. We initially independently classify each segmented tooth of a dental film based on teeth reconstruction in four image subspaces established using Principal Component Analysis (PCA). Then we validate and possibly correct initial teeth class assignments using teeth - neighborhood rules. Finally, if the resulting teeth class sequence is unique we assign a number for each tooth corresponding to its position in the dental quadrant it belongs to. Otherwise, if the resulting teeth class sequence is non-unique we call a reject option.

Testing results based on a sizable dataset suggest that our low-cost features achieve fast classification with an average film classification time of .8 seconds. Overall around 75% of the teeth are initially assigned to their proper classes, 86% of the teeth are assigned proper class labels after validation, and around 70% of the teeth assigned proper intra-quadrant numbers. The Accuracy of classification and numbering in bitewing films are approximately 5% higher than the overall results. By adjusting the penalization scheme for misclassification, we could achieve a teeth-numbering accuracy of 82% in bitewing films, with classification accuracy of 97% for premolars, and 89% for molars, also 93% of teeth were numbered within 1 neighbor with respect to the ground truth.

Chapter 4: Micro and Macro Decision Making

Dental radiographic records usually contain multiple films, which in turn capture multiple views of teeth. In order to reach more reliable decisions on the match status between a subject case and a potential match case, we exploit two aspects of teeth multiplicity.

In this chapter we present a hierarchical fusion scheme, which at its lower level fuses the multiple match scores, due to the multiple views of a given tooth-pair, into a single match score for that tooth-pair, and hence produce a *micro* decision on the match status of that tooth pair. At the top level of our fusion scheme we aggregate the different micro decisions into a single *macro* decision on the match status between the subject case and its potential match counterpart.

4.1 Overview

Dental records used in postmortem identification are usually composed of multiple radiographic films, each of which shows a number of teeth (e.g. *see* Figure 21). Often, some teeth appear in multiple films with multiple views depending on the film type (e.g. periapical, bitewing, panoramic) and the imaging setup, i.e. the relative positions of the x-Ray tube, the film and the teeth.

When comparing multiple views of a subject tooth to multiple views of the corresponding tooth of a potential match (or candidate), we obtain as many match scores as the number of unique pairings of the views of the subject tooth and the reference tooth. So, for example if the upper left first molar shows in three films of a subject record and in two films of a candidate record, then we obtain six, most likely different, match scores.

The sources of variability in these match scores are possibly the different details shown by the multiple views of a tooth as well as the possible transformations and occlusions caused by imperfections in the various preprocessing steps, e.g. teeth segmentation, and teeth alignment. To reach a decision on whether a subject tooth matches a reference tooth, it is important to combine the different match scores, resulting from the different view pairings, into a single match score, which we then use as basis for making this *micro* decision.

We refer to the tooth-pair match decision as a micro decision to distinguish it from the case-to-case match decision, which we refer to a *macro* decision. Macro decision-making is based on aggregation of the micro decisions, and it thus fully exploits teeth multiplicity.

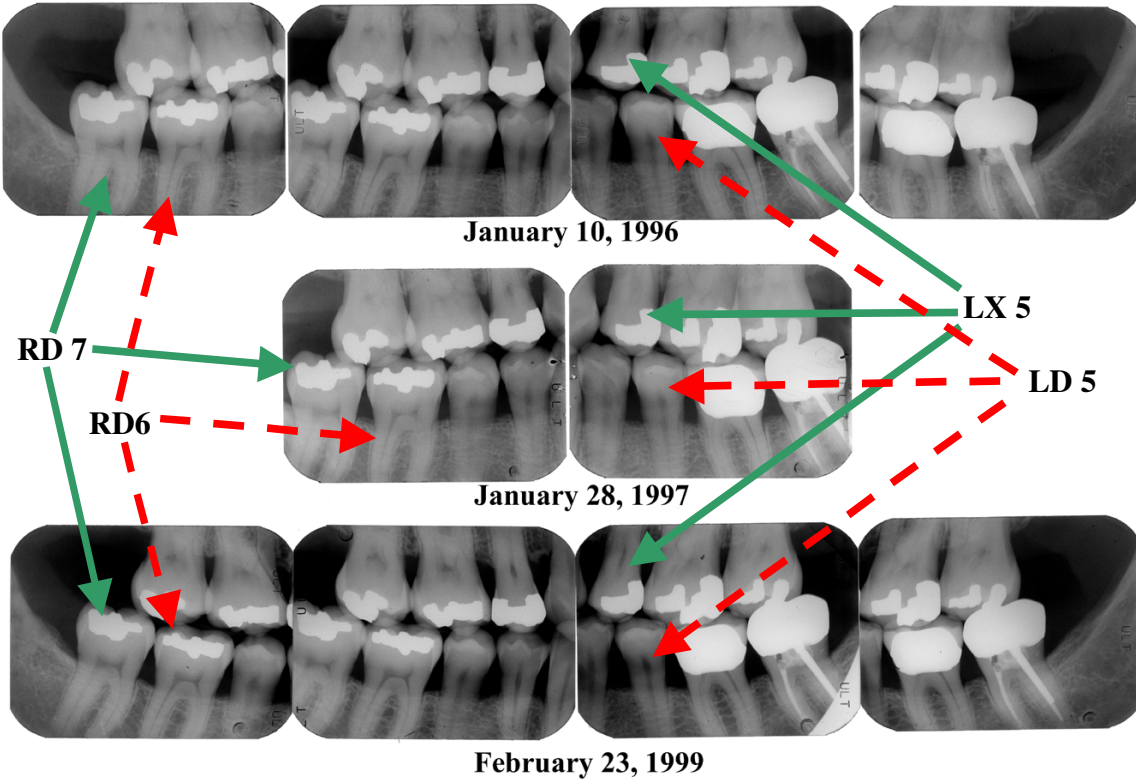


Figure 21: A dental record with multiple films showing multiple views of teeth.

To tackle the problems of micro and macro decision-making we propose a hierarchical fusion scheme, which exploits view multiplicity of teeth, as depicted in Figure 22: The bottom level of the fusion scheme addresses the question of combining match scores resulting from comparison of multiple views of the same subject/reference tooth-pair into a single probability of match between that subject/reference tooth-pair and accordingly generating micro decisions. The top level of the fusion scheme answers the question of combining the micro decisions due to the different teeth-pairs in order to produce a macro decision on the match status between the subject/reference case-pair.

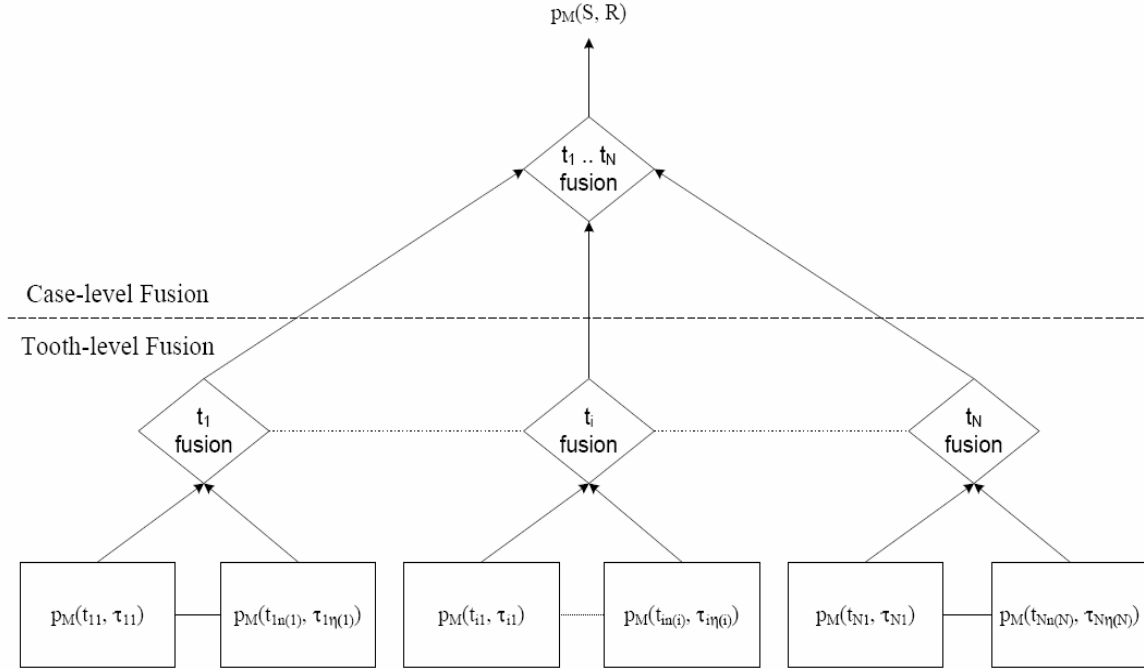


Figure 22: A hierarchical fusion scheme for micro and macro decision-making.

In section 4.2 we present our approach for tooth-level for micro decision-making. In section 4.3 we describe how we fuse micro decisions into a case-level decision (or macro decision), then we introduce a ranking scheme for sorting matched cases in a match list. Finally in section 4.4 we present our experimental results.

4.2 Micro-Decision Making

Because a tooth may appear in multiple films of a dental record, we think that this view multiplicity should be exploited in making a more reliable decision on whether a subject/reference tooth-pair is matched. In earlier work we addressed the problem of matching one view of a subject tooth to one view of a reference tooth, where we estimated a probability of match (or match score) between the pair given that view.

At the tooth-level fusion, we want to aggregate the probabilities of match between the different views of tooth t_i of a subject case $(t_{i1}, \dots, t_{in(i)})$ and the available views of tooth τ_i of a reference case $(\tau_{i1}, \dots, \tau_{in(i)})$, such that we produce a single number representing the probability of match between the subject tooth t_i and reference tooth τ_i , namely $p_M(t_i, \tau_i)$. We

want to also make a decision (micro decision), namely $d_M(t_i, \tau_i)$, on whether tooth t_i matches tooth τ_i .

We think that the different views of a tooth should not be considered independent observations, and accordingly we rule out the product rule as a possible option for fusion at the tooth-level fusion. We study three alternative rules for fusing the posterior match probabilities (match scores), namely the mean rule, the median rule, and the maximum rule. Thus given multiple match scores due to the different views of t_i and τ_i , i.e. $\{p_M(t_{ik}, \tau_{il}); k = 1, \dots, n(i), l = 1, \dots, \eta(i)\}$, the fused match score ($p_M(t_i, \tau_i)$) according to each of these three rules read:

$$\begin{aligned} \text{Mean Rule} & : p_{MMEAN}(t_i, \tau_i) = \frac{1}{n(i)\eta(i)} \sum_{\langle k, l \rangle} \{p_M(t_{ik}, \tau_{il}); k=1, \dots, n(i), l=1, \dots, \eta(i)\}. \\ \text{Median Rule} & : p_{MMED}(t_i, \tau_i) = \text{med}_{\langle k, l \rangle} \{p_M(t_{ik}, \tau_{il}); k=1, \dots, n(i), l=1, \dots, \eta(i)\}. \\ \text{Maximum Rule} & : p_{MMAX}(t_i, \tau_i) = \max_{\langle k, l \rangle} \{p_M(t_{ik}, \tau_{il}); k=1, \dots, n(i), l=1, \dots, \eta(i)\}. \end{aligned}$$

Once we determine the tooth-to-tooth match score ($p_M(t_i, \tau_i)$) according to any of the match score fusion rules, we compare it to two thresholds to determine whether we think the evidence is sufficient to declare the tooth pair unmatched or matched, or we think that the evidence is not sufficient to make a firm decision, and hence declare the micro decision as undetermined. The rule we use for hardening $p_M(t_i, \tau_i)$ to a micro decision $d_M(t_i, \tau_i)$ is [67]:

$$d_M(t_i, \tau_i) = \begin{cases} \text{UnMatched} & \text{if } p_M(t_i, \tau_i) \leq 0.2 \\ \text{Matched} & \text{if } p_M(t_i, \tau_i) \geq 0.8 \\ \text{Undetermined} & \text{otherwise} \end{cases}$$

4.3 Macro-Decision Making and Ranking

With N teeth in common between the dental charts of a subject case and a candidate case, we obtain N micro decisions ($N \leq 32$). At the case-level fusion, we are looking for a scheme for combining these N micro decisions $\{d_M(t_i, \tau_i), i = 1, \dots, N\}$ into a macro decision $D_M(S, R)$ between that subject case S and its potential match counterpart R .

In macro decision-making we are fusing decisions and hence the only fair and suitable fusion scheme is the majority-voting rule. Thus with N micro-decisions $\{d_M(t_i, \tau_i), i = 1, \dots, N\}$, the majority voting rule reads: $D_M(S, R) = \mathcal{Q}_j \mid j = \arg \max_v \{N_v\}; v \in \{1, 2, 3\}$, $\mathcal{Q}_j \in \{\text{'Matched'}, \text{'Unmatched'}, \text{'Undetermined'}\}$. Where N_1 , N_2 , and N_3 respectively indicate the number of instances where $d_M = \text{'Matched'}$, $d_M = \text{'Unmatched'}$, and $d_M = \text{'Undetermined'}$ such that $N = N_1 + N_2 + N_3$.

We also use N_1 , N_2 , and N_3 to compute a rank score $\rho_M(S, R)$, which helps us in sorting the match list. We think of the rank score $\rho_M(S, R)$ as function $g(N_1, N_2, N_3)$ with the following desirable characteristics:

- g is non-decreasing in both N_1 and N_3 . So, as either the number of micro matches and/or the number of undetermined micro decisions increases, $\rho_M(S, R)$ should not decrease.
- g is non-increasing in N_2 . Conversely, as the number of the micro mismatches increases, $\rho_M(S, R)$ should not increase.
- $g(32, 0, 0) = 1$. As ultimately for a subject/reference pair that has 32 matched teeth (the maximum number of teeth is a normal adult), this reference record should be examined before any others that appear in the match list.
- $g(0, N_2, 0) = 0$. As the function g should be grounded for $N_1 = N_3 = 0$. Moreover, this corresponds to a record that will not be placed in the match list to begin with.
- $g(0, 0, 32) = 1/2$ (by rational choice).

One possibility for the ranking function g is $g_1(N_1, N_2, N_3) = \frac{(\sqrt{2} N_1 + N_3)^2}{64(N_1 + N_2 + N_3)}$. Figure 23

shows the profile of the proposed function $g_1(N_1, N_2, N_3)$, Figure 24, Figure 25, and Figure 26 show slices of $g_1(N_1, N_2, N_3)$ at $N_1 = 0$, $N_2 = 0$, and $N_3 = 0$ respectively.

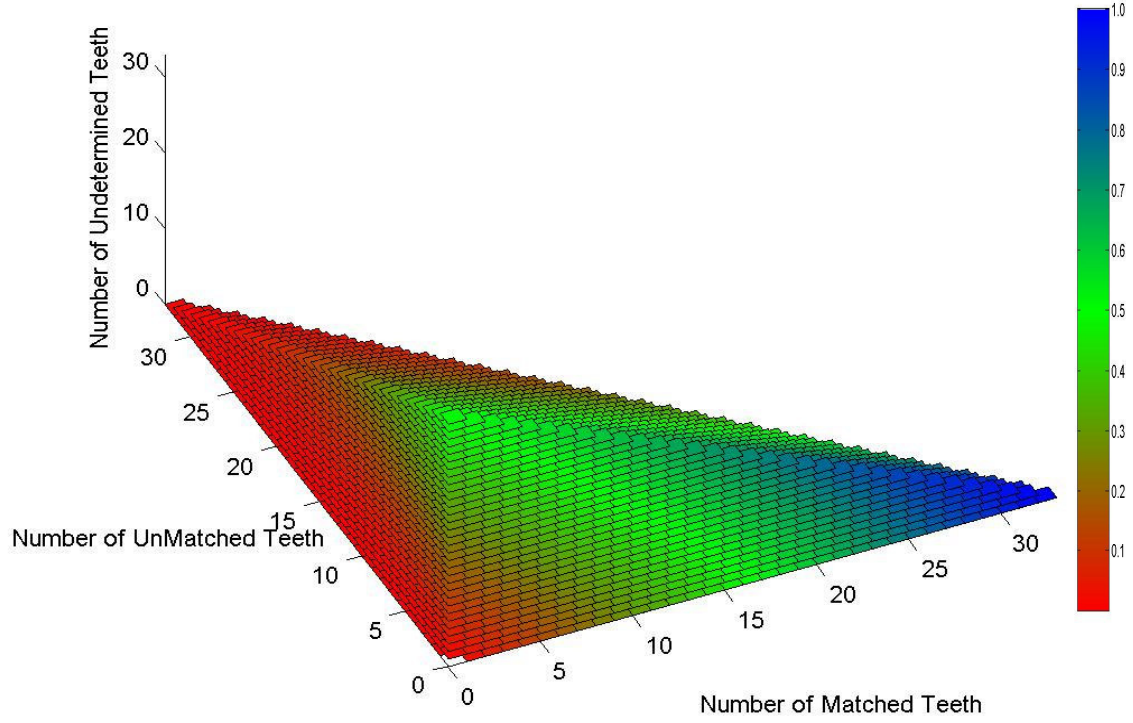


Figure 23: Profile of the function $g_I(N_1, N_2, N_3) = \frac{(\sqrt{2} N_1 + N_3)^2}{64(N_1 + N_2 + N_3)}$

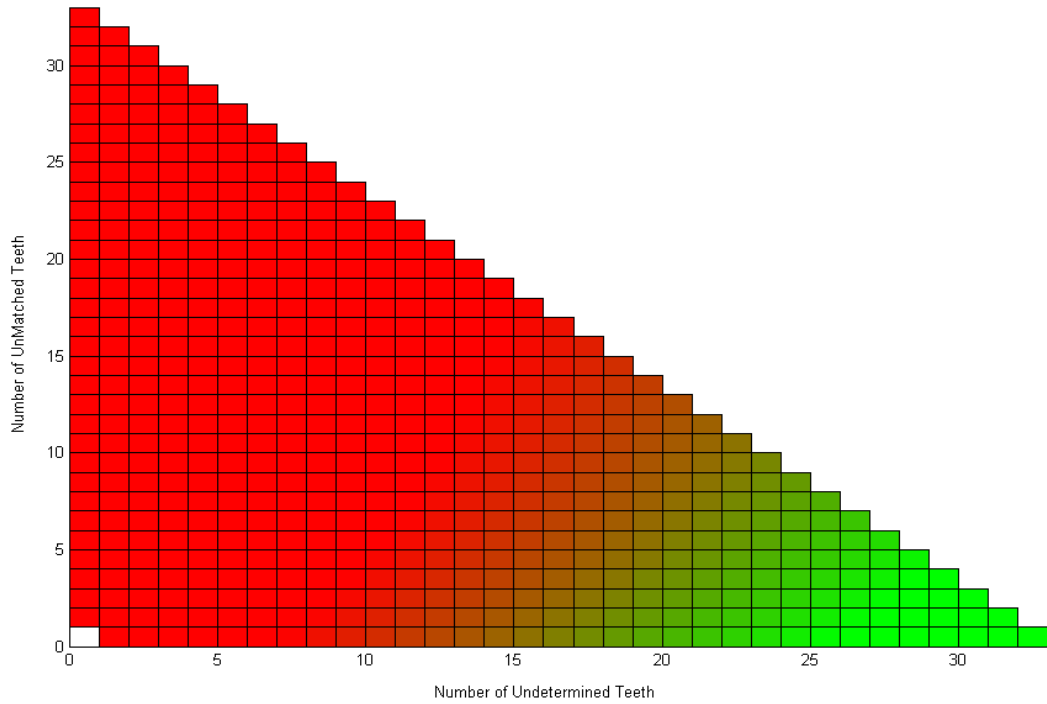


Figure 24: Profile of the function $g_I(0, N_2, N_3) = \frac{N_3^2}{64(N_2 + N_3)}$

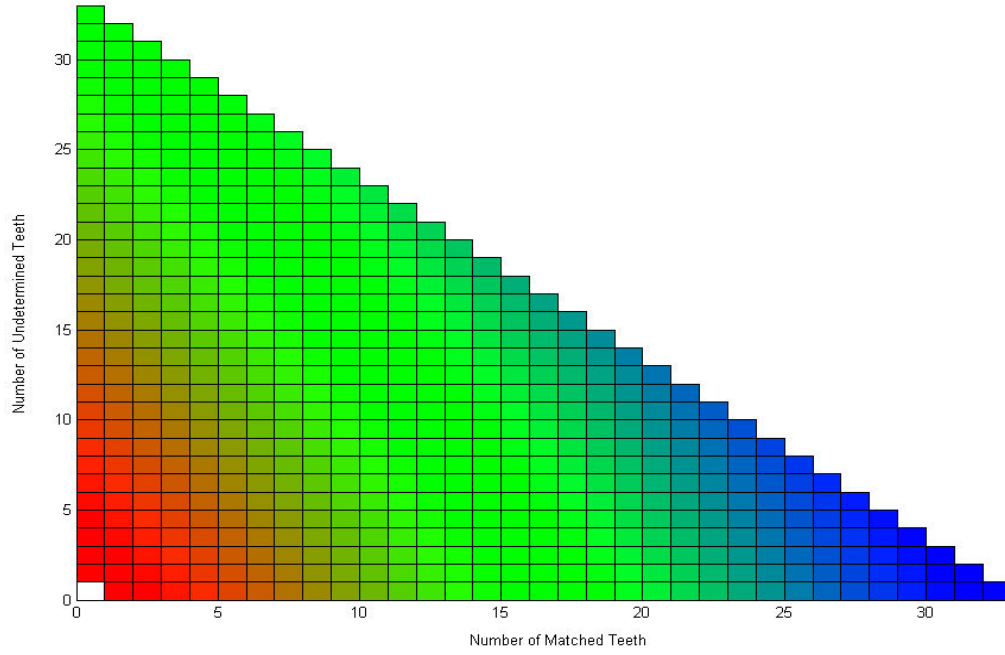


Figure 25: Profile of the function $g_I(N_I, 0, N_3) = \frac{(\sqrt{2} N_1 + N_3)^2}{64(N_1 + N_3)}$

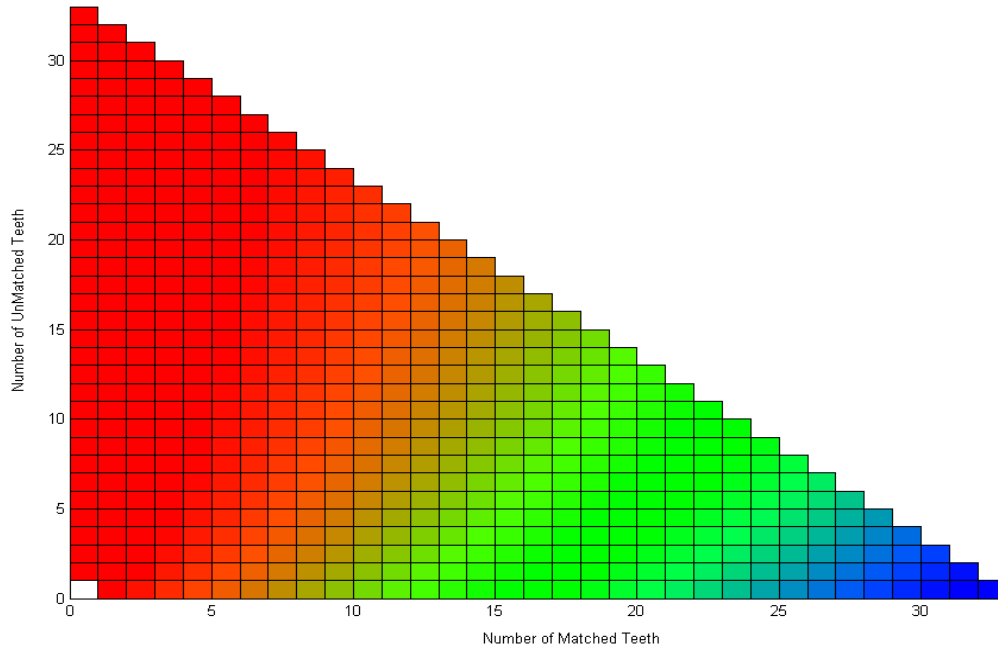


Figure 26: Profile of the function $g_I(N_I, N_2, 0) = \frac{N_1^2}{32(N_1 + N_2)}$

An alternative choice for g is $g_2(N_1, N_2, N_3) = \frac{(\sqrt{2} N_1 + N_3)^2}{2(N_1 + N_2 + N_3)^2}$, in which we normalize

the ranking scores with respect to the number of teeth in common between the subject and each of the references, i.e. $N_1 + N_2 + N_3$. This in contrast to division by 32 in the case of g_1 , which is the absolute maximum value for $N_1 + N_2 + N_3$.

4.4 Experimental Results

We conducted identification experiments using a dataset of 106 cases extracted from the CJIS dental image databases [64][65] and the test set prepared by Dr. Robert Howell [66]. The number of the test cases labeled as AM is 48 test cases, and 58 test cases are labeled as PM cases, with 23 AM/PM matched pairs. We preprocessed these cases using the record cropping technique described in [8], the film enhancement technique described in [9], the segmentation techniques described in [9] and [10] and the teeth labeling technique described in Chapter 6. We inspected the preprocessing data of each of these cases and excluded erratic data.

We divided the experiments in two sets. In the first set we used the 23 AM cases, which have PM matches, as subjects. We formed a candidate list for each subject from among the 58 PM reference records such that a candidate shares at least four teeth with the subject, we used the dental charts of the subject and the PM records to determine the number of shared teeth between a subject and a reference. In the second set we followed the same procedure but using the 23 PM cases, which have AM matches, as subjects and we used the 48 AM cases as references.

The total number of case pairs that we compared is 397 pairs, 46 out of which are matched. The total number of teeth-pair views that we compared is 9900 out of which only 1368 are actually matched. Figure 27 shows the distribution of the teeth-pair views involved in our experiments; we notice that premolars and molars dominate the chart, because these teeth appear in dental films more often. The results we present in this section are all based on uncompiled MATLAB® realizations running on a 512 MB RAM, 2.8 GHz Intel P4® based personal computer.

We used the MR-GA approach [18][20] for tooth-to-tooth alignment but assuming the quasi-affine model suffices for correcting geometric misalignments between two views of a tooth. While this assumption is not accurate (*see* [37]), however the alignment approaches presented in [16] and [13] assume an even simpler model. Hence we thought that reducing the dimension of the search space from 6D (in [18] and [20]) to 5D would cut on the time of the GA search at the cost of ignoring misalignments due to shear. We limited the alignment time to 30 seconds per tooth-pair view.

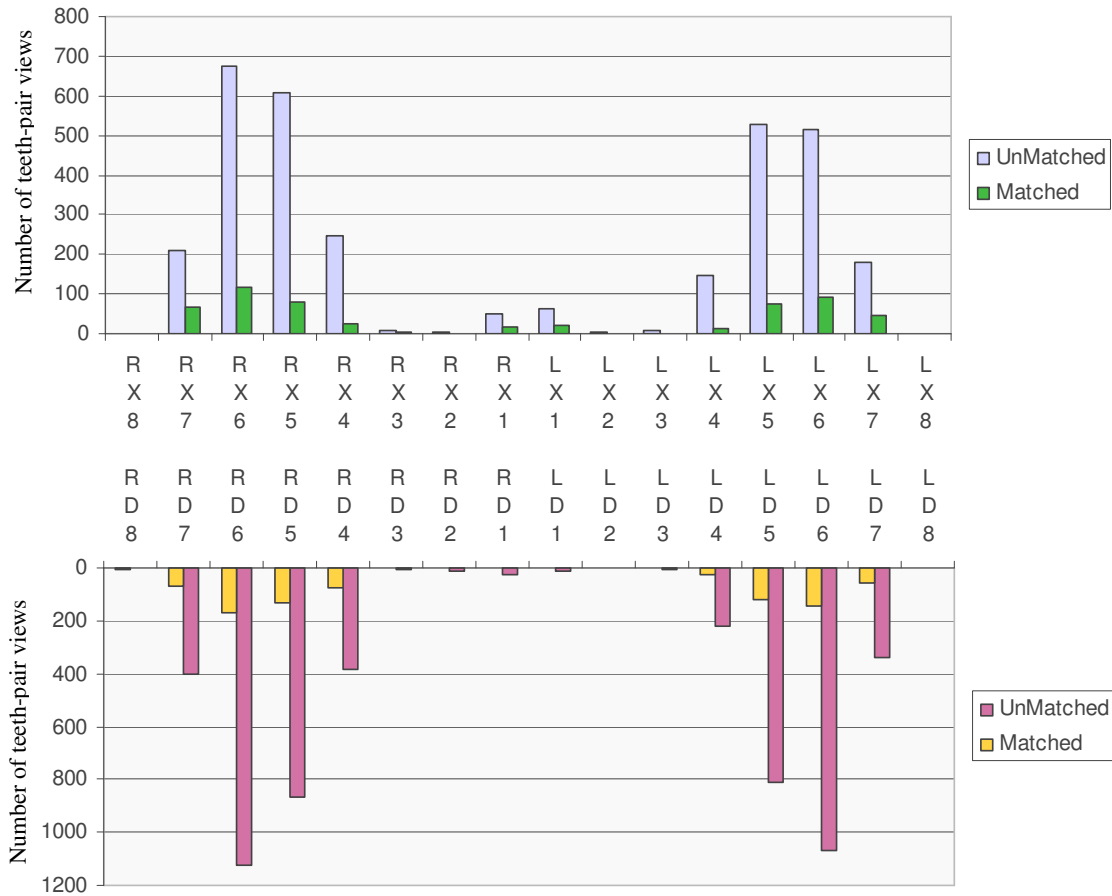


Figure 27: Distribution of the teeth-pair views of the test dataset.

The total elapsed time for the experiment was approximately 100 hours, predominantly spent in alignment, with a mean case-to-case matching time of 15 minutes, with about 80% of the case-pair comparisons completed within 20 minutes as inferred from Figure 28, which shows the distribution of the case-to-case matching time for the experiments reported herein. The case-to-case matching time largely depends on the number of common teeth between the

case-pair, as well as the number of different views of a subject/reference tooth-pair. In one case, the total number of view comparisons between a subject and a candidate was 126.

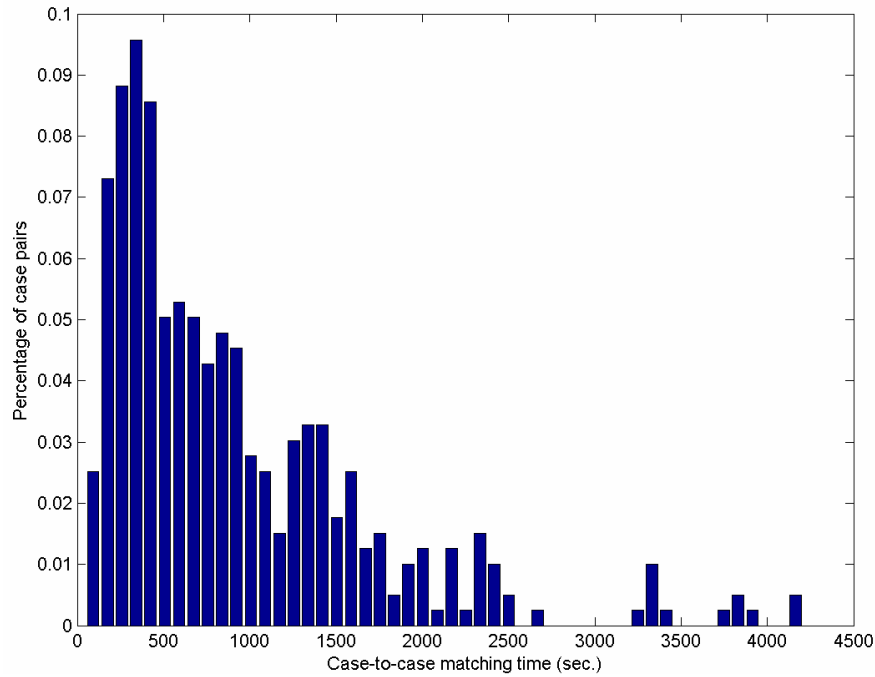


Figure 28: Distribution of case-to-case matching time.

Table 6 lists the error rates for genuine matches as observed at the macro-decision level. We notice that of the three alternative tooth-level fusion schemes, the max fusion rule led to the highest genuine accept rate ($\sim 87\%$) and the lowest genuine reject rate ($\sim 13\%$). However, if we consider the undetermined cases as cases that will not be dropped from the final presentation of results to the forensic expert, then the mean and median rules perform reasonably close to the max rule.

Table 6: Error rates for genuine matches at the macro decision level.

	Mean rule	Median rule	Max rule
Genuine Accept Rate	30.43 %	47.83 %	86.96%
Genuine Undetermined Rate	54.35 %	28.26 %	0
Genuine Reject Rate	15.22 %	23.91 %	13.04 %
Imposter Accept Rate	26.50 %	23.08 %	30.77 %

The error rates presented in this section are based on the two fixed thresholds of 0.2 and 0.8 for micro decision making that we selected based on an earlier study [67]. In future studies we shall vary these thresholds and examine their effect on macro decisions.

Figure 29 depicts the hit rate for case-to-case matching based on the ranking function

$$g_1(N_1, N_2, N_3) = \frac{(\sqrt{2} N_1 + N_3)^2}{64(N_1 + N_2 + N_3)} \text{ for each of the three tooth-level fusion schemes, while}$$

the hit rate for case-to-case matching based on the ranking function $g_2(N_1, N_2, N_3) =$

$$\frac{(\sqrt{2} N_1 + N_3)^2}{2(N_1 + N_2 + N_3)^2} \text{ is shown in Figure 30.}$$

When the top four or less records of the match list are considered, we observe that normalizing the ranking score using the absolute maximum number of common teeth between subjects and references (i.e. 32 teeth), which is the case of $g_1(N_1, N_2, N_3)$, produces better hit rate compared to normalizing it with respect to the actual number of common teeth between the subjects and each of the references separately. However, when five or more of the top matches are considered the chart due to $g_2(N_1, N_2, N_3)$ is superior to that due to $g_1(N_1, N_2, N_3)$.

Both hit-rate charts shown in Figure 29 and Figure 30 suggest that while the max rule results in the highest genuine accept rate and the lowest genuine reject rate, the more conservative rules (the mean and the median rules) outperform the more optimistic rule (the max rule) in the sense of ranking the match list.

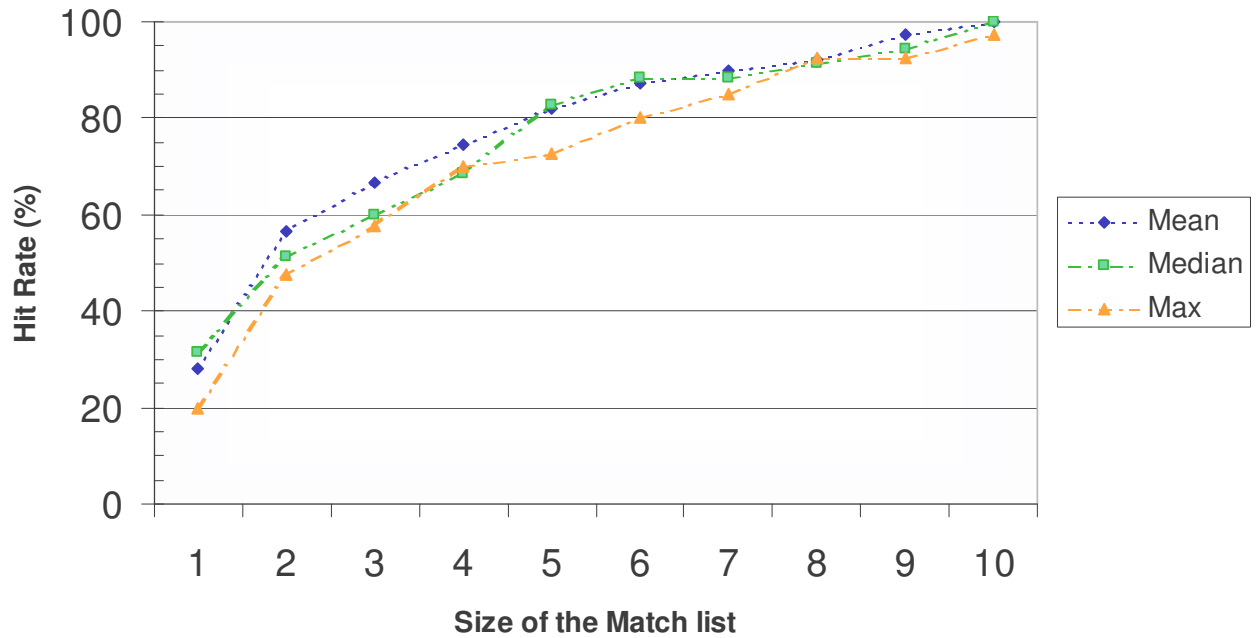


Figure 29: Hit rate based on fixed normalization of the ranking score.

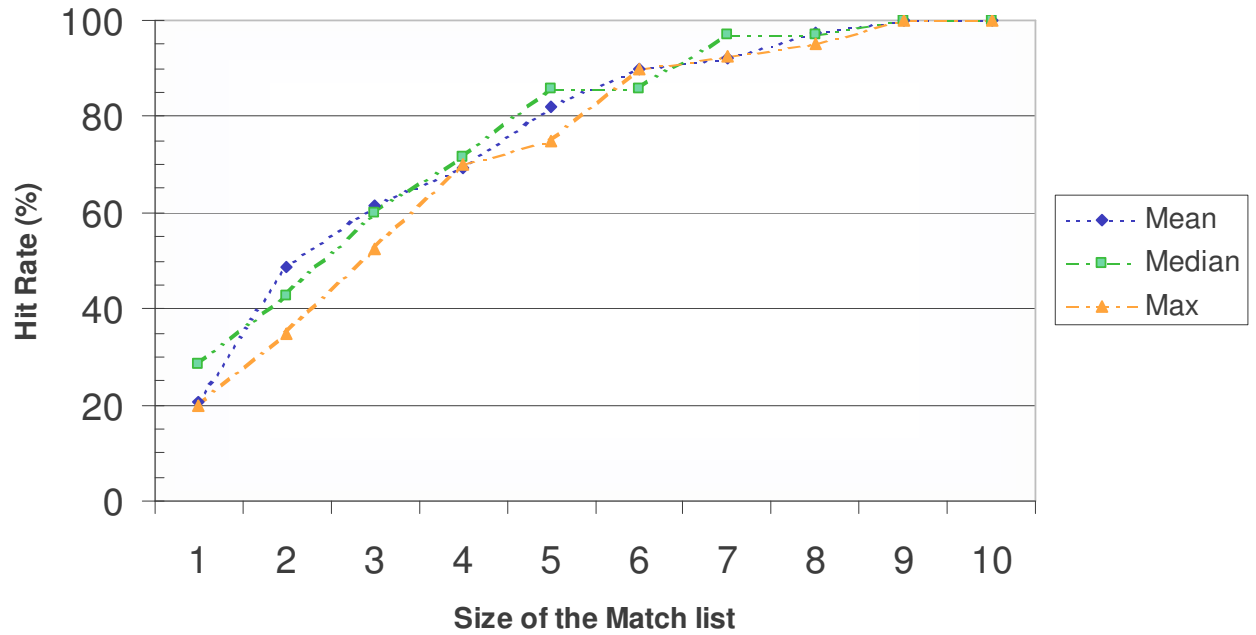


Figure 30: Hit rate based on variable normalization of the ranking score.

In Figure 31 (a, c, e) we show examples of subject records and we show the records that were ranked top in the match lists of these records in Figure 31 (b, d, f) respectively. The records shown on Figure 31 (b, d) are the genuine matches to those shown in Figure 31 (a, c), whereas the record shown in Figure 31 (f) is a false match to that shown in Figure 31 (e).

Note that the case-pair shown in Figure 31 (a, b) suggest our approach posses some robustness that counts for flipped films.

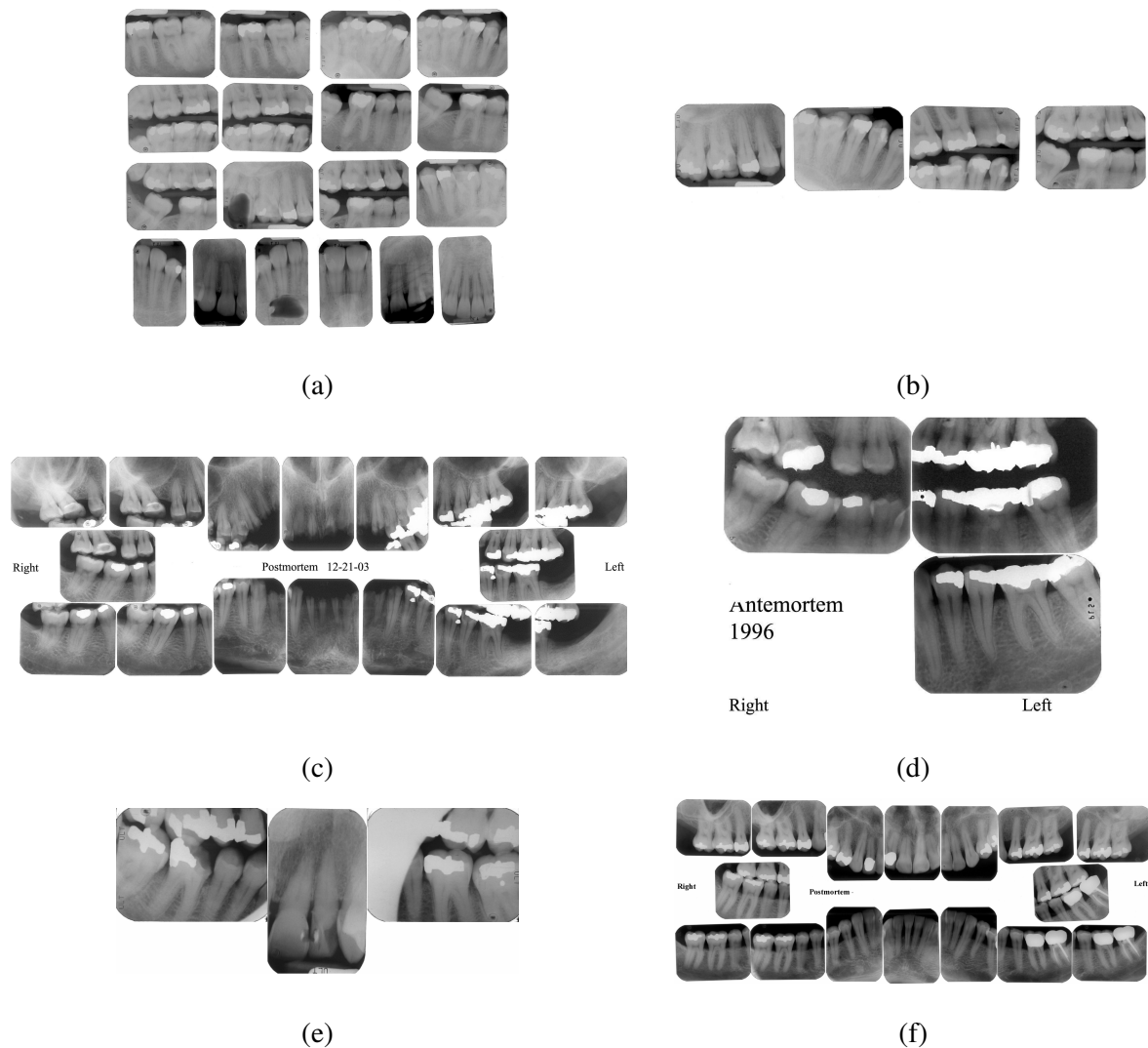


Figure 31: Examples of subject records (a and c) and their respective matches (b and d), ranked first in the match lists of these subjects, and a subject record (e) with a false accept record (f) ranked top of the match list of that subject.

4.5 Summary

Dental radiographic records usually contain multiple films, which in turn show multiple views of teeth. To achieve reliable match decisions between a pair of cases based on dental

characteristics, we devised a hierarchical fusion scheme that exploits two aspects of teeth multiplicity.

The lower level of our fusion scheme addresses the problem of combining multiple match scores of a tooth-pair due to the possible multiple views of the pair in the different radiographic films, and hence establishing a micro decision on the match status between a subject/reference tooth-pair.

The upper level of the hierarchical fusion scheme aggregates the micro decisions due to the different subject/reference teeth pairs to produce a macro decision on the match status between the subject record and the reference record. We also developed a scheme for ranking the different possible matches of a subject in a match list.

We instantiated three alternatives for micro decision-making based on the mean, median, and maximum rules for match score fusion. As macro decision-making relies on hardened micro decisions, we resorted to the majority-voting rule for case-to-case match decisions.

Testing results based on 397 case pairs, which involved comparison of 9900 teeth-pair views, show that the case-to-case genuine reject rate is around 15% based on the mean and the max rules. The more conservative mean rule showed superiority in terms of higher hit rates near the top of the match lists. The mean rule achieves the best compromise between false accept rate and false reject rate.

Chapter 5: On Performance Limits of Dental Identification

The performance of a biometric system depends, among other factors, on the capability of the feature representation scheme to capture the intra-subject invariance and the inter-subject variance, ideally retained by the biometric signal. Biometric features should desirably possess high complexity (*randomness*) so that they suffice for discriminating between individuals of a large population [68].

In this chapter we study the problem of the limitations of performance of dental identification due to adopting a set of features. We approach this problem following well-known approaches in the literature, and then we propose a generalized approach to determining the number of degrees of freedom, as a figure of merit in the study of performance limits. We also look at a workable approximation of our generalized approach.

5.1 Overview

Accuracy aspects (or *error rates*) set the hard performance limits for a biometric identification system [57]. To meet certain accuracy requirements, a biometric system should accommodate two types of variability: (i) *within-subject* (or *intra-class*) variability, and (ii) *between-subject* (or *inter-class*) variability [57][58][59]. It is generally desirable that a biometric identifier minimizes the first type of variability, thus minimizing the *false-reject rate* and maximizes the second type of variability, thus minimizing the *false accept rate* [59].

The factors that limit the accuracy of a biometric identification system are its: (i) “*information-limited behavior*”, (ii) “*representation-limited behavior*”, and (iii) “*invariance-limited behavior*” [57]. The first class of factors pertains to the biometric signal acquisition process, where the biometric phenomenon, the environment and the sensor characteristics set a *lower error bound* (or a *higher performance bound*). The second class of accuracy-limiting factors stems from the imperfections of the feature representation scheme, which may not retain all the useful information (intra-class invariance and inter-class variations), and thus *raises* the error bound above (or *lowers* the performance bound below) that due to the inherent information-limitation of the biometric signal (its *source entropy* [68]). The last class of factors tends to further *raise* the error bound by failing to perfectly absorb the *intra-class variance* of all subjects.

Forensic scientists have used dental characteristics for decades as a means for positive postmortem identification [1]. However, the dental identification has only been recently tackled from a biometrics angle, i.e. seeking to achieve a highly automated system for dental identification. A few dental features were proposed as basis for dental identification within the research framework of the Automated Dental Identification System (ADIS). For example in [16], *Chen* and *Jain* use teeth contours as templates and they also use shapes of dental works. In [13] *Nomair* and *Abdel-Mottaleb* use salient points of teeth contour to generate invariant teeth signatures. In [69] *Nassar* and *Ammar* use appearance-based features of teeth learnt using a neural network.

While the research on automating the process of dental identification is fairly recent, it is interesting to analyze the proposed dental features and examine the performance limits of dental identification owing to using specific features. Thus, it becomes important to decouple the overall performance of an identification system from its performance limits due to the features it uses. This decoupling is particularly important in understanding the root cause(s) of performance limitations. For example, if a given biometric system does not achieve a desirable performance, it may be the matcher models that need to be readdressed and not necessarily the set of features.

The objective of the study we present in this chapter is to estimate the limits (expectations) on identification performance due to two types of dental features: First, templates of teeth contours, and second, our appearance-based features used in the image comparison component of the prototype ADIS (*see* Figure 2 and the Appendix *page 104*).

In section 5.2 we study the limits on the performance of identification based on teeth contours, we tackle this problem using a similar approach to that of the individuality of fingerprints proposed by *Pankanti et. al.* in [60]. In section 5.3 we study the limits on the performance of dental identification using our appearance-based features. We first tackle this problem using the approach proposed by *Daugman* in [59] and [68] by estimating the number of degrees of freedom possessed by the features, with the assumption of identical distribution of the independent feature dimensions. Then we introduce a generalized approach for determining the number of degrees of freedom based on the joint entropy of a vector random variable, as a figure of merit in determining the information capacity of the feature set.

Finally, we work out a second order approximation that use *copulas* [70][71] and *mutual information* [72][73] to identify dependencies between the feature dimensions and hence help in obtaining a better estimate of the number of degrees of freedom.

5.2 Performance Limits of Dental Identification using Teeth Contours

We tackle the problem of performance of identification based on teeth contours using a combinatorial feature-matching approach similar to that proposed in [11][12]. We make the following assumptions:

- Teeth contours extracted from radiographic images are used as features for identification of individuals by matching contour points of a subject against those of references (or candidates) [11]. The reference contour points are designated $P_R = \{(x_R(j), y_R(j)), j=1..m\}$, and the subject contour points are designated $P_S = \{(x_S(i), y_S(i)), i=1..n\}$.
- Each tooth fully fits in a rectangular bounding box of width w pixels and height h pixels as shown in Figure 32, and the contour of a tooth fully fits in this rectangular bounding box. A tooth contour is assumed to be of length l and while different teeth images may have different scanning resolutions, in addition to the fact that different teeth may naturally have different contour lengths, we think that this assumption is reasonable as we may always rescale points such that the contour length assumption is valid.
- A subject contour point $P_S(i)$ is associated with (i.e. marked as a *unique* match of) a reference contour point $P_R(j)$ based on the following *matching distance* criterion:

$$\text{decide: } \begin{cases} \text{Matched} & \text{if } d_M(p_S(i), p_C(j)) \leq r_o, \\ \text{UnMatched} & \text{otherwise} \end{cases},$$

where $d_M(p_S(i), p_R(j)) = \sqrt{(x_S(i) - x_R(j))^2 + (y_S(i) - y_R(j))^2}$ is the *Euclidean* distance between points $P_S(i)$ and $P_R(j)$ in a common *Cartesian* coordinate system, r_o is a match distance threshold, which is empirically estimated and $d_M(p_S(i), p_R(j)) = \min_k \{d_M(p_S(k), p_R(j))\}, k = 1..n$, with n being the total number of the subject contour points.

- A *rigid-body* geometric transformation model (allowing for rotation, uniform scaling, and translation) [34] suffices to describe view variations between two images of one tooth, which is the same assumption *Chen* and *Jain* make in [11].
- In the course of matching a pair of teeth contours, there is an *effective* area of overlap (A) between the two images, and because matching is based on the objects' peripheries, not all points bounded by this area are equally interesting. In other words, the *effective* area (A) is rather confined around the contours.
- Each contour point is surrounded by a *tolerance area* (C); any point, which belongs to another set of contour points, and falls within this area is a potential match to the first point, however recall that only one point-pair association is permissible. This tolerance region is important for possible misalignments.

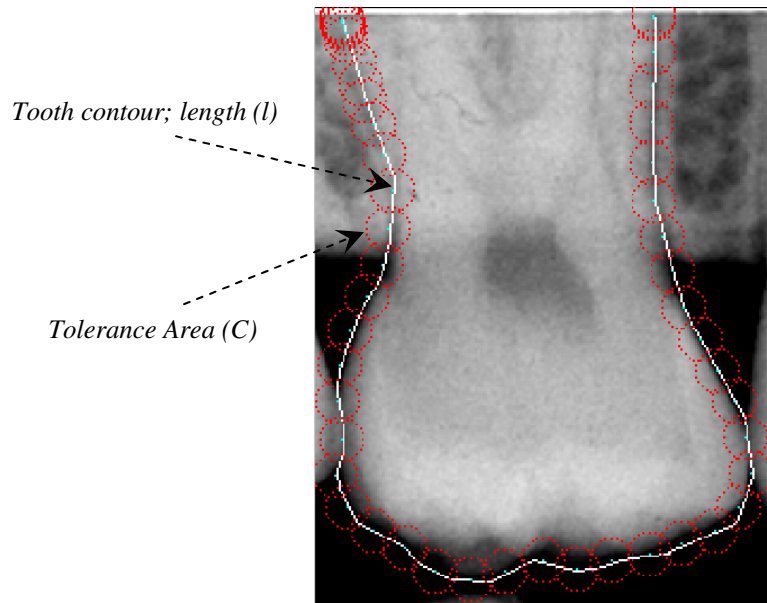


Figure 32: A rectangular box bounding a tooth, also tolerance areas around its contour points

In order to determine the probability of false association between a subject and a reference tooth, let us consider the following scenarios:

- Given a reference tooth contour represented by m -points and a contour of a subject tooth with a single point in the overlap area A , the probability of matching that point of the subject contour to any of the m -points of the reference contour is the

probability that the subject point falls in any of the tolerance areas surrounding the m -reference contour points, which is: $\frac{mC}{A}$.

- Given an m -point reference contour and 2-points of the subject contour that fall in the overlap area A , the probability of matching *exactly* one of the two subject contour points to any of the m points of the reference contour is thus: $2 \cdot \frac{mC}{A} \cdot \frac{(A-mC)}{(A-C)}$, the factor of 2 is because any of the 2-subject points may be the point that matches any of the reference m points, the $\frac{(A-mC)}{(A-C)}$ factor is the probability that the point designated as *unmatched* to any of the m points has to fall else where in $(A-C)$ where the m reference points exist.
- Similarly, given an m point reference contour and n -points of the subject contour that fall in the overlap area A , the probability of matching *exactly* one of the n subject contour points to any of the m -points of the reference contour is: $\binom{n}{1} \cdot \frac{mC}{A} \cdot \frac{(A-mC)}{(A-C)}$
- Eventually, given an m -point reference contour and n points of the subject contour that fall in the overlap area A , the probability of matching *exactly* q of the n subject contour points to q points of the m -points of the reference contour is:

$$p(q | A, C, m, n) = \binom{n}{q} \cdot \left[\frac{mC}{A} \cdot \frac{(m-1)C}{A-C} \cdot \dots \cdot \frac{(m-q+1)C}{(A-(q-1)C)} \right] \times \left[\frac{(A-mC)}{(A-qC)} \cdot \frac{(A-(m-1)C)}{(A-(q+1)C)} \cdot \dots \cdot \frac{(A-(m-(n-q+1))C)}{(A-(n-1)C)} \right]$$

Because typically $A \gg C$, we define $M = \left\lceil \frac{A}{C} \right\rceil$, where M represents the total number of *non-overlapping* tolerance areas within the overlap area A . We may accordingly approximate the probability $p(q | A, C, m, n)$ to $p(q | M, m, n)$ as follows:

$$\begin{aligned}
p(q | M, m, n) &\approx \binom{n}{q} \cdot \left[\frac{m}{M} \cdot \frac{m-1}{M-1} \cdots \frac{m-q+1}{M-(q-1)} \right] \times \\
&\quad \left[\frac{M-m}{M-q} \cdot \frac{M-(m-1)}{M-q-1} \cdots \frac{M-(m-n+q-1)}{M-(n-1)} \right] \\
&= \binom{n}{q} \cdot \frac{m!}{(m-q)!} \cdot \frac{(M-n)!}{M!} \cdot \frac{(M-m)!}{((M-m)-(n-q))!} \\
&= \binom{m}{q} \binom{M-m}{n-q} / \binom{M}{n}
\end{aligned}$$

The resulting probability model is the *hyper-geometric* distribution of random variable q with parameters M, m, n . The underlying experiment is independent withdrawal without replacement of m and n samples from two identical pools of M -distinct samples, where the *hyper-geometric* distribution gives the probability that out of the independently withdrawn m and n samples, q samples are identical (note that $q \leq \min(m, n)$). Thinking in terms of error probability in biometrics, this distribution models the chance that q out of M feature containers randomly filled with m and n binary events from S and R where $\max(m, n) \leq M$, are doubly occupied by features from S and R , given that each of S and R may not “double-fill” any of the M containers. We observe that similar to fingerprint minutiae; contour points are also binary events.

5.2.1 Parameter Estimation

The approach we followed for determining performance limits of dental identification using teeth contours requires two parameters (A, C), which are necessary to determine the value of M . Once M is known, we estimate performance limits of dental identification based on teeth contour matching. In this study, we will only focus on the parameters A and C .

We first start by estimating C , the tolerance area around contour points. As this is a proof of concept study, we have used a small dataset of 15 matched teeth-pairs (antemortem vs. postmortem), which belong to 6 different cases in the CJIS dental image database [65]. We have manually extracted contours for these teeth-pairs as shown by the example in Figure 33.

We observe that dense sampling of contours heavily violates, in principle, independence between features as well as potentially oppose the assumption of single occupancy (per identity) of the hypothetical M feature-containers. Therefore, we subsample the reference

contours such that the general contour shape is preserved (we chose to keep 50 points, such that they maintain some fair circular distribution as well as a near fixed segment lengths).

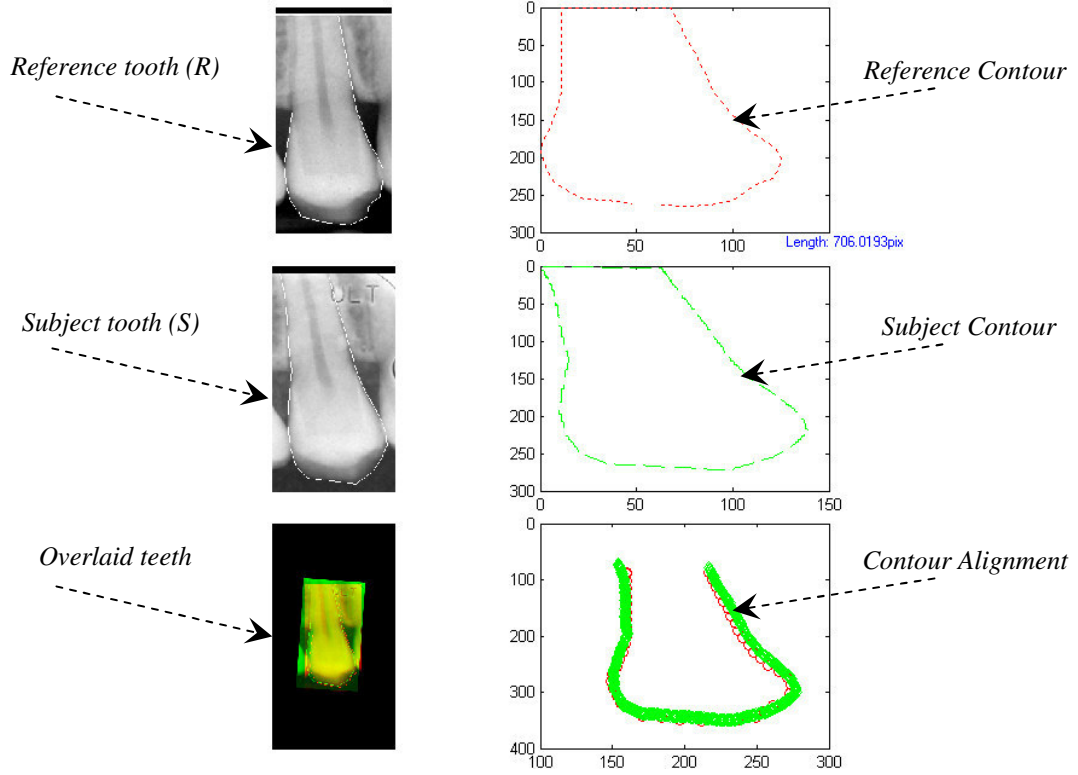


Figure 33: Example of manual matching of teeth-pairs using manually extracted contours, contour points are not plotted to scale.

We manually aligned contour-pairs (and hence teeth-pairs), and computed the matching distance using the distance maps of the transformed subject contours (we kept the reference contours (teeth) fixed and transformed the subjects). On registration we rejected outlier reference contour points whose distance from the closest point on the transformed subject contour is unusually large ($> 5 \cdot D_{av}$, where D_{av} is the average match distance between the reference points and the transformed subject contour). Figure 34 depicts the cumulative probability distribution (based on the histogram of point matching distance for 641 matched point-pairs). Using, we may conclude that for 99% of the matched point-pairs in this study $d_M(p_S(i), p_R(j)) \leq 10 \text{ pixels}$. Hence, we safely set the *matching distance threshold* $r_o = 10 \text{ pixels}$.

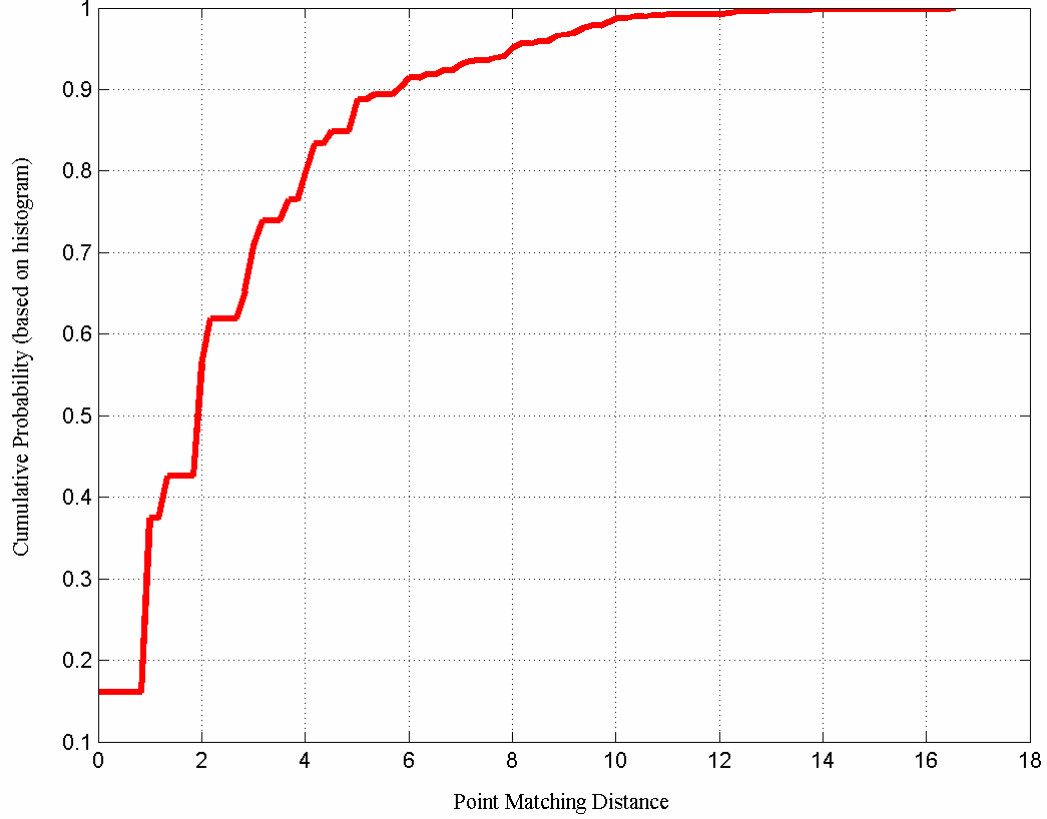


Figure 34: Cumulative probability distribution of point matching distance using 641 points.

We have also computed the length of reference contours used in this experiment and found that the average length was 795.8 pixels and the median length was 822.7 pixels. As we mentioned, we think the *effective* overlap area is confined around the contour (as shown in Figure 32), thus we may approximate A as $A \approx 2r_o l$. Also the tolerance area may be

reasonably set to $C = \pi r_o^2$. Hence, $M = \left\lceil \frac{A}{C} \right\rceil = \left\lceil \frac{2l}{\pi r_o} \right\rceil$. Using the median contour length, our estimate for M is $M = 53$ (distinct feature-containers).

5.3 Performance of Dental Identification using appearance-based features

In this section we present a preliminary attempt to study the limits on the performance of dental identification using our appearance-based features. We tackle this problem from two viewpoints: In section 5.3.1, we use the approach proposed by *Daugman* in [59] and [68] based on determining the number of degrees of freedom possessed by the features, under the

assumption of identical distribution of the independent feature dimensions. In section 5.3.2 we introduce a generalized approach to the problem using *copulas* [70][71] and *mutual information* [72][73] to identify dependencies between feature dimensions and hence determine the actual information capacity of the feature set. And finally in section 5.3.3, we work out first and second order approximations to estimate the number of degrees of freedom as a figure of merit in determining the performance limits.

The appearance features described in [6][21][69] and in the *Appendix* (see page 104), where a tooth region of interest (*ROI*) is compressed into a 32×32 image, leading a *multivariate* discrete random variable, each of its dimensions requires up to 8-bits for representation. We extract two feature images from each *ROI*, namely Z^1_{ROI} and Z^2_{ROI} , each of which is a 6×6 feature image (or a 36-dimensional feature vector). Because each dimension corresponds to processing a 7×7 receptive field from the *ROI*, each dimension may thus require up to $8 \times 7 \times 7$ bits for its representation. Therefore each of the feature images is a 6×6 *multivariate discrete* random variable, with each of their dimensions requiring up to 392 bits for representation.

5.3.1 Estimating Limits on Performance under IID feature dimensions

In [59] and [68] *Daugman* uses a statistical hypothesis testing approach to address the question on the uniqueness of an IrisCode[®]. An IrisCode[®] is a 256 Byte binary sequence that describes textural features extracted from iris images [59][68]. While the absolute upper bound on the capacity of the IrisCode[®] suggests that up to 2^{2048} different individuals can be represented by this scheme, however, the actual population size that can be uniquely represented by this scheme should be much lower. This is due to the correlations (or more precisely dependencies) between the bits of the IrisCode[®] [68].

There are two sources for redundancy in a feature representation of a given biometric, the first is inherent to the biometric signal and the second is due to the filtering operations performed to obtain the feature representation [68]. *Daugman* [59][68] establishes that correlations across the dimensions of a biometric feature representation, as well as correlations across the population of that representation of a biometric feature, both tend to lower the upper limit on the information capacity of the feature representation scheme and hence raises the lower bound on identification error.

To estimate the lower error bounds delimited by features in the case of dental identification, based on our appearance-based features, we pursue the statistical testing approach proposed by *Daugman*. From a practical point of view we need not consider all 2^{392} possible values that may be assumed by each of the 72 dimensions of our appearance-based dental feature vector. On one hand, it is infeasible to obtain a sample size that is sufficient for studying the distribution of a 28,224-bit feature vector. And on the other hand, if we recall our interpretation of the feature filters (*see* Appendix section A5.2.c), one of the two kernels seems to extract gradient information from 7×7 tiles of the image. Thus, the values produced by this kernel inform us about directions of intensity gradients in some sense. Over such a small tile area, it makes more sense to think of only a few gradient directions. Hence, we quantize the dental features such that each dimension is represented by 4 bits, with 16 equally separated quantization levels per dimension. Accordingly, we are seeking to determine the number of degrees of freedom in a 36 Byte *dental* code.

We designate the number of degrees of freedom estimated using samples of unmatched pairs as N_U , and similarly that estimated using samples of matched teeth pairs as N_M . These numbers help us model some distance distributions describing the variability in both matched and unmatched pairs, and hence determine the limits on performance due to features.

Given a binary representation of the features sample extracted from paired and aligned teeth, we determine the distribution of the average *Hamming* distance [72] between code-pairs. The average Hamming distance between two code words (A and B) is given by

$$\overline{D}_H = \frac{1}{L} \sum_{\langle i \rangle} A_i \oplus B_i, \text{ where } L \text{ is the length of each code word, } A_i \text{ and } B_i \text{ are the } i^{\text{th}} \text{ bit of the}$$

first and second code words respectively, and \oplus designates the logical exclusive OR operation. The range of the average Hamming distance is $[0, 1]$ and ideally the average Hamming distances between codes corresponding to matched pairs should generally be smaller than those corresponding to unmatched pairs.

Assuming that all bits of the binary representation are independent from each other but identically distributed (which is an inaccurate assumption particularly in our case), then the underlying experiment would be that of N independent and identical Bernoulli trials. Thus, leading to a binomial distribution, whose number of degrees of freedom is given by

$N = \frac{\mu(1-\mu)}{\sigma^2}$, where μ and σ are the mean and standard deviation computed from the distribution of the average Hamming distance \bar{D}_H .

We conducted experiments to determine the number of degrees of freedom in the cases of premolars and molars. We used feature samples of automatically aligned pairs of premolars (100 fairly aligned matched pairs and 1500 unmatched pairs) and pairs of molars (100 fairly aligned matched pairs and 1500 unmatched pairs). Figure 35 (a) shows the distribution of the average Hamming distance between premolar codes, and Figure 35 (b) shows that distribution between molar codes. In Table 7 we list the statistics measured from the samples of the average Hamming distance.

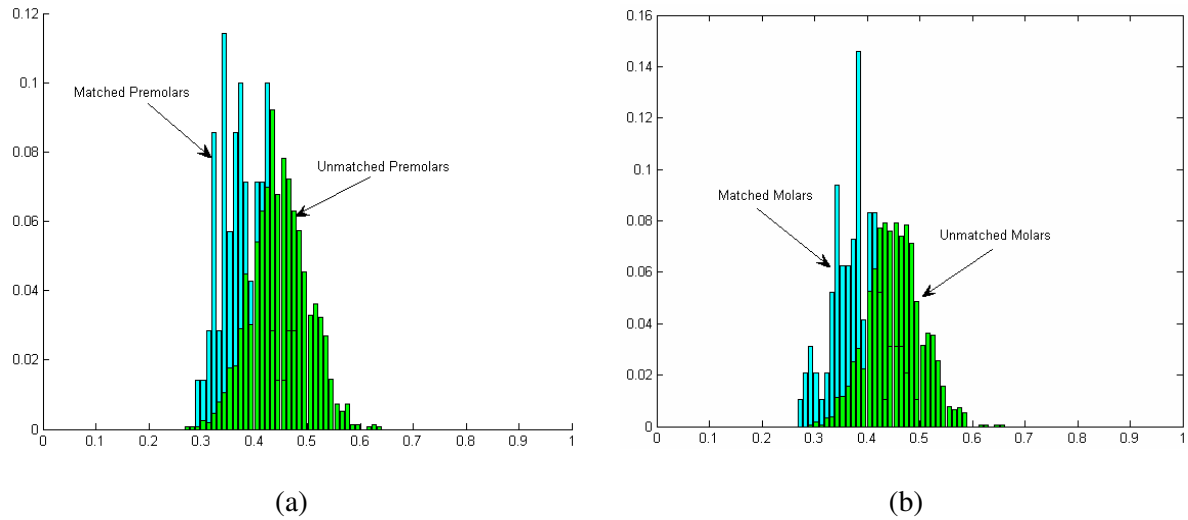


Figure 35: Average Hamming distance distribution for (a) Premolars, and (b) Molars.

Table 7: Sample statistics for the average Hamming distance of premolars and molars.

	Matched	Unmatched
μ	0.38051	0.44846
σ	0.04409	0.05332
N	121	87

(a) Premolars

	Matched	Unmatched
μ	0.38111	0.45385
σ	0.04734	0.05152
N	105	93

(b) Molars

We observe that the distributions of the average Hamming distance in the cases of genuine and imposter teeth are not well separated primarily due to the intra-subject variability. The

decidability index ($d' = \frac{|\mu_{matched} - \mu_{unmatched}|}{\sqrt{(\sigma_{matched}^2 + \sigma_{unmatched}^2)/2}}$ [58][59]) for molars is 1.47 and for

Premolars is 1.39. By fitting binomial distributions to the parameters we obtained for Premolars and Molars, as shown in Figure 36, we determine theoretical FAR and FRR associated with different Hamming distance thresholds as listed in Table 8. Equal error rates are achieved at a threshold of 0.412 for Premolars and a threshold of 0.416 for Molars.

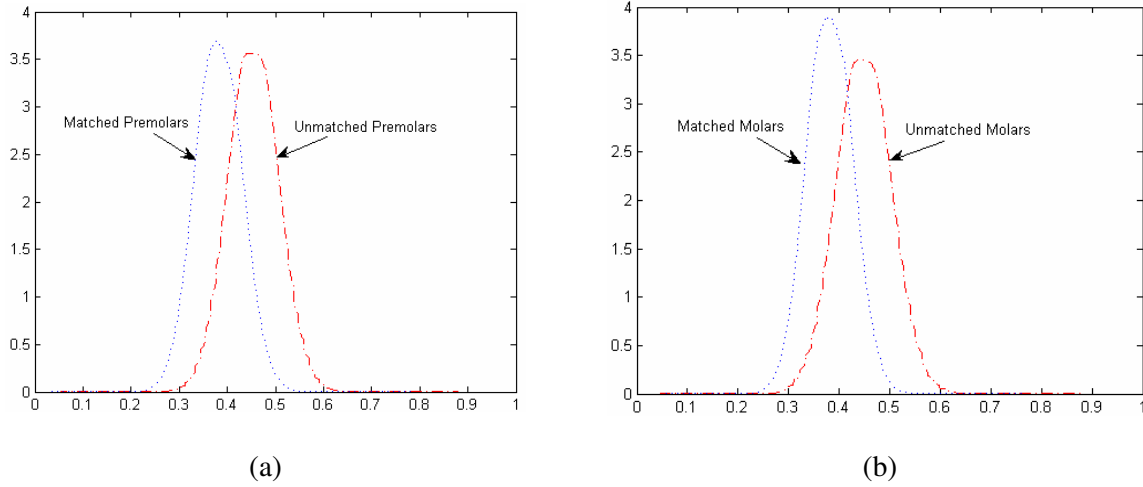


Figure 36: Binomial fitting of the average Hamming distance distributions for (a) Premolars, and (b) Molars.

While the FAR and FRR figures are atypical of standard biometrics, we recall that teeth multiplicity is a distinctive characteristic of dental identification. Thus, with multiple teeth compared, the probability of making errors in all compared pairs exponentially decreases.

Table 8: Variation of performance limits with decision thresholds.

Threshold	FAR	FRR
0.400	0.179	0.331
0.410	0.234	0.254
0.412	0.243	0.243
0.420	0.296	0.189
0.430	0.364	0.135
0.438	0.420	0.100
0.440	0.436	0.092
0.450	0.510	0.060
0.460	0.583	0.038

(a) Premolars

Threshold	FAR	FRR
0.400	0.147	0.351
0.410	0.195	0.276
0.416	0.232	0.232
0.420	0.253	0.210
0.430	0.319	0.155
0.440	0.392	0.110
0.443	0.412	0.100
0.450	0.469	0.076
0.460	0.546	0.050

(b) Molars

5.3.2 Estimating the Number of Degrees of Freedom: A Generalized Approach

Our estimate of the number of degrees of freedom in section 5.3.1 is based on the assumption that the independent bits are identically distributed. While in the case of IrisCode[®] this

assumption may be valid to some good extent, with other biometric features this assumption is not very accurate, e.g. our appearance-based dental features. In this section we offer a generalized approach to estimating the number of degrees of freedom under a less stringent assumption on the “*identity*” of the independent feature bits. The number of degrees of freedom of a sequence of Bernoulli experiments is the effective number of bits needed for the representation of that sequence, i.e. its entropy. If the experiments underlying the sequence were identically distributed, the resulting distribution would be a binomial distribution. However, if these underlying experiments are not identically distributed, the resulting distribution is not the binomial distribution.

If we imagine flipping N coins with different biases, then the probability of getting m heads

$p(m, N)$ is given by: $p(m, N) = \sum_{i=1}^K \prod_{j=1}^m p_{s(i,j)} \prod_{k=1}^{N-m} (1 - p_{t(i,j)})$, where:

- $K = \binom{N}{m}$
- $s(i, j)$ designates a grouping function that indicates the positions m of N bits that are set under the (i, j) grouping pair,
- $t(i, j)$ designates a grouping function that indicates the positions $N-m$ of N bits that are not set under the (i, j) grouping pair,
- $s(i, j)$ and $t(i, j)$ have the following properties:
 - $\forall (i, j), s(i, j) \cap t(i, j) = \Phi$
 - $s(i_1, j_1) \cap s(i_2, j_2) = s(i_1, j_1) \Rightarrow i_1 = i_2 \wedge j_1 = j_2$
 - $t(i_1, j_1) \cap t(i_2, j_2) = t(i_1, j_1) \Rightarrow i_1 = i_2 \wedge j_1 = j_2$

We ultimately want to compute the joint entropy of the bit sequence in order to reach a better estimate of the number of degrees of freedom N , without resorting to the strong assumption of i.i.d bits. Copulas [70][71] are special functions that link the marginal distributions of multiple random variables to a multivariate joint distribution. The methodology we propose for determining the number of degrees of freedom is outlined in the following steps:

- We use feature samples of matched and unmatched pairs to obtain binary representation of the features,
- We detect the dissimilarity between the respective binary sequences in the two cases of features due to matched pairs as well as features due to unmatched pairs. We achieve this using the bit-wise XOR logical operation,
- We estimate the marginal distribution of the feature differences along the different dimensions, by:
 - Segmenting the bit sequences due to the XOR operation into words, each word corresponds to a feature dimension,
 - Converting the binary words into quantization levels,
 - Computing the histograms of the quantization levels for each of the dimensions.
- We determine the joint distribution with knowledge of the marginal distributions using a multivariate copula function, and
- We compute the entropy of the resulting joint distribution, which is our estimate of the degrees of freedom.

Evidently our choice of a copula affects the estimate of the degrees of freedom, and this choice should either be made based on an educated guess or by trying several copulas and choosing the one that achieves the minimum discrepancy between the sample joint distribution and the predicted joint. The value of using copulas in this methodology is to establish a model for the inter-dimension dependence.

5.3.3 Estimating the Number of Degrees of Freedom: Approximations

While the methodology that we presented in the previous section on estimating the number of degrees of freedom constitutes a general approach to the problem, it may not be practical to study multivariate joint distributions. In this section we look at first order and second order approximations of the joint entropy of a multivariate distribution.

Let:

- \underline{Z} be a d dimensional vector random variable,
- N_{z_i} denote the number of bits representing z_i (the i^{th} dimension of \underline{Z}),
- $H(z_i)$ denote the entropy of z_i ,
- $H(z_{r_1(i)}, z_{r_2(j)})$ denote the joint entropy of two different dimensions of \underline{Z} , namely

$z_{r_1(i)}, z_{r_2(j)}$, where $\forall (i, j) \in \{1, 2, \dots, d/2\} \times \{1, 2, \dots, d/2\}$:

- $i \neq j \Rightarrow r_1(i) \neq r_1(j)$,
- $i \neq j \Rightarrow r_2(i) \neq r_2(j)$,
- $r_1(i) \neq r_2(j)$

To estimate the number of degrees of freedom, N , we want to compute the joint entropy

$$H(\underline{Z}), \text{ we observe that } H(\underline{Z}) \leq \sum_{i=1}^{d/2} H(z_{r_1(i)}, z_{r_2(i)}) \leq \sum_{i=1}^d H(z_i) \leq \sum_{i=1}^d N_{z_i}.$$

The zeroth order approximation of $H(\underline{Z})$ is simply the actual number of bits used in the binary representation of \underline{Z} , which is $\sum_{i=1}^d N_{z_i}$, this approximation assumes that all bits are

independent and are equally distributed. The first order approximation of $H(\underline{Z})$ surpasses the uniform distribution assumption, however, still does not depart from the assumption of bit independence. The first order approximation is given by $\sum_{i=1}^d H(z_i)$, i.e. the sum of the entropies of the different dimensions of \underline{Z} .

The second order approximation is what we are more interested in, as the previous approximations pertain to trivial situations. This approximation is based on computing the joint entropies due to pairing of dimensions according to $\sum_{i=1}^{d/2} H(z_{r_1(i)}, z_{r_2(i)})$. To achieve a

good second order approximation, we should pair-up the dimensions that lead the least sum of the joint entropies of the pairs. Given that we can determine the joint distribution for any two pairs, as we show later, the strategy we follow for choosing the pairs to be used for the second order approximation is as follows:

- We compute the mutual information between all possible pairs of the feature dimensions,
- We sort the pairs in descending order according to their mutual information,
- We select that top $d/2$ pairs $z_{r_1(i)}, z_{r_2(j)}$, where $\forall (i, j) \in \{1, 2, \dots, d/2\} \times \{1, 2, \dots, d/2\}$:
 - $i \neq j \Rightarrow r_1(i) \neq r_1(j)$,
 - $i \neq j \Rightarrow r_2(i) \neq r_2(j)$,
 - $r_1(i) \neq r_2(j)$.

This pairing strategy does not necessarily guarantee the least sum of the joint entropies of the pairs, however, it is a structured strategy to seek a reasonable second order approximation. One of the steps incorporated in determining the second order approximation of the joint entropy is to estimate the joint distribution of a pair of dimensions, which we achieve by letting:

- $u = F(z_i)$, and $v = G(z_j)$ denote the cumulative marginal distributions of z_i and z_j respectively,
- $J_\psi(z_i, z_j)$ denote the cumulative joint distribution of z_i and z_j estimated using a feature sample ψ ,
- $J_{C_k}(z_i, z_j)$ denote the cumulative joint distribution of z_i and z_j estimated using u, v and a copula C_k , where we limit the copulas considered in this study to:

- $C_1(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}$, $-1 \leq \theta < 1$, which is the *Ali-Mikhail-Haq* copula family with single parameter (θ) [70],
- $C_2(u, v) = uv + \gamma uv(1-u)(1-v)$, $-1 \leq \gamma \leq 1$, which is the *Farlie-Gumbel-Morgenstern* copula family with single parameter (γ) [70],

We can show that the value of θ that minimizes the discrepancy between $J_\psi(z_i, z_j)$ and $J_{C_1}(z_i, z_j)$ in a least square error sense is given by:

$$\theta = \frac{\sum_{\langle i \rangle \langle j \rangle} (u_i v_j - J_{\psi}(z_i, z_j)) u_i v_j (1 - u_i)(1 - v_j)}{\sum_{\langle i \rangle \langle j \rangle} (1 - u_i)^2 (1 - v_j)^2 u_i v_j J_{\psi}(z_i, z_j)}$$

We can also show that the value of γ that minimizes the discrepancy between $J_{\psi}(z_i, z_j)$ and $J_{C_2}(z_i, z_j)$ in a least square error sense is given by:

$$\gamma = \frac{\sum_{\langle i \rangle \langle j \rangle} (u_i v_j - J_{\psi}(z_i, z_j)) u_i v_j (1 - u_i)(1 - v_j)}{\sum_{\langle i \rangle \langle j \rangle} (u_i v_j (1 - u_i)(1 - v_j))^2}$$

By applying the proposed methodology for approximating the joint entropy, using a sample of 1500 unmatched pairs of Premolars and 1500 unmatched pairs of Molars, we found the second order approximation of the joint entropy in the case of Premolars to be 248 bits and that in the case of Molars to be 197 bits. The first order approximations assuming independence of the features dimensions gives 260 bits in the case of Premolars and 236 bits in the case of Molars.

5.4 Alternative Representations

The studies we conducted in sections 5.2 and 5.3 consider performance of dental identification using single tooth evidence. With teeth multiplicity being one of the distinctive aspects of dental identification it is interesting to consider dental features that depart from the tooth level to multiple teeth; especially that identification performance based on features of a single tooth does not compare to the performance of identification in other biometrics. In this section we sketch a scheme for multiple teeth representation.

We first consider dental features based on teeth contours. Suppose we are given a series of films that show multiple teeth, and suppose we have a reliable way for extracting teeth contours from these films. We reduce the contour of each tooth to a few number of salient points, and represent each tooth contour by a sequence defining the length and orientation of the line segments connecting each two consecutive points in a given cyclic order, i.e. similar to the idea of the chain codes for boundary representation [74] but without the hard restrictions on segment lengths and orientations. Other alternative representations based on

teeth contours include identifying a set of more complex shape primitives and expressing the contours in terms of operations on these primitive. Salient points can be contour points with interesting geometric properties, e.g. high curvature points and inflation points, or anatomically significant points, e.g. the cemento-enamel junction.

To accommodate the variability of teeth representations under uniform scaling, we normalize the lengths of the line segments between the salient points of a tooth contour by dividing these lengths by their sum. The next step is to generate a sequence that describes the lengths and orientations of the line segments. By studying the distributions of the lengths and differential angles, we can quantize these two entities and define an alphabet of normalized lengths and differential angles. Hence, obtain a code describing each tooth.

The next level in this representation scheme achieves description of multiple teeth by concatenating the descriptors of each contour after specifying the relation, also as connector length and orientation, between each two consecutive reference points (e.g. shape centroids). Using special characters, we can specify the classes of teeth, the types of salient points, and the extent of the sequence describing the shape of each tooth. Matching of two cases thus becomes a problem of string matching with error. We think that such a representation, if it can practically be realized, is ideal for addressing the potential match search problem (*see* Figure 1).

Given that this teeth representation scheme can be realized, then the question on the performance limits of dental identification using these features can be addressed by the same approach we presented in section 5.3.

We attempted several approaches for automatic contour extraction, however, none of the approaches we pursued led to accurate contours. The problem of reliable automatic extraction of teeth contours remains an unsolved challenge.

The representation of multiple teeth using the appearance-based features is straightforward; we concatenate the features. Hence, the performance limits of dental identification using appearance features of multiple teeth reduces to following the same approach presented in section 5.3 but using a longer feature vector. Due to the unavailability of a sizeable database of genuinely matched teeth at the time of conducting this study, we defer the analysis of performance limits of dental identification to future work.

5.5 Discussion

In the study we presented in section 5.2, we may think of the contour representation of a tooth as a *single-ridge* fingerprint, however, the *tooth minutiae* (contour points) possess less randomness than do fingerprint minutiae, this is primarily due to the *confinement* of the former to a smaller area (around the contour), and while the match distance threshold we estimated in this study is lower than the *15 pixel* distance threshold in [60], the smaller overlap area in case of teeth contours predominates, and thus M in the case of dental contours is smaller than that of a fingerprint.

As the effective overlap area A between the reference and the subject image is concentrated around the contour of the former, all feature points on the latter which lie on A are considered matched, and thus $n = q$. In which case, the probability of false acceptance according to the approach we pursued in section 5.2 reduces to $p(n|M, m) = \frac{\binom{m}{n}}{\binom{M}{n}}$.

The denominator of this distribution is the number of ways n -points of the subject contour may be placed in the M -containers, the numerator is the number of ways n points of that contour may be associated with the m reference points.

One may easily use the distribution $p(n|M=53, m)$ to determine the *similarity-limited upper bound* [57] on the probability of *false acceptance* when two contours, one of which represented by m points, have exactly n matched points (according to the match distance criterion). So, for example there is a false association probability of about *0.0003* when we make a dental identification match decision based on matching 12 points of a subject tooth contour to a candidate tooth contour defined by 30 points. Recall, that one of the assumptions made in developing this model relate to feature independence, which may not hold under some circumstances, and accordingly practical probabilities of *false acceptance* are likely to be much higher.

While it seems intuitive that it would always be desirable to make dental match decisions based on as many contour points as possible, dental contours inherently bear limitations in their identification capacity due to the relatively small number of *binary* feature containers. When we attempt to fit more of these binary features into the limited number of containers (M), we are *implicitly* forcing a match decision and consequently incur a higher probability

of *false matches*. For example when we conduct dental identification by matching 41 points of a subject contour to a 52-point reference contour the probability that this be a *false match* is 0.22642.

One way around lowering the false match probability is to further exploit contour points by considering other aspects (e.g. orientation) not only the locations of the contour points, and thus the M containers will not only be filled by *binary* features. However, reliable automatic extraction of teeth contours is a major problem, which remains an unsolved challenge.

We also think that confining the representation of dental images to teeth contours tend to miss important other features, like the shapes of teeth pulp chambers, teeth bone-density profiles and of course dental work, which may be considered a biometric in a broad sense. We believe that all these features are also important for improving the performance of dental identification.

In the study presented in section 5.3, we may think of 72 feature containers filled by colored features. The study we presented tried to statistically determine the expected discrepancy in colors due to instances of matched and unmatched teeth. While we bear in mind that alignment imperfections partly contribute to raising the color discrepancies in the case of matched teeth; we also recall that dental characteristics, especially when appearance-features are considered, constitute an atypical biometric in the sense that possible physical transformations of teeth are enormous, e.g. restorations, fractures, extraction, etc. Hence the intra-subject variance may be comparable to the inter-subject variance in many cases. We think that it is thus important to use as many teeth as possible to exclude potential mismatches.

5.6 Summary

The performance of a biometric system depends, among other factors, on the capability of the feature representation scheme to capture the intra-subject invariance and the inter-subject variance, ideally retained by the biometric signal.

In this chapter we determined bounds on dental identification based on the capabilities of two sets of dental features, the first uses teeth contours, and the second is appearance-based. We analyzed the performance limits due to the first set of features using a counting approach to

determine the probability of false co-occurrence, and we analyzed the performance of the second set of features using a coding approach.

We offered a generalized methodology that used copulas to determine the number of degrees of freedom, as a figure of merit in performance analysis, to relax the stringent assumptions on feature independence when studying performance limits. We also presented workable approximations of the proposed approach.

Chapter 6: Conclusions and Future Work

This chapter summarizes the outcome of the research presented in this dissertation and sheds some light on directions for possible future extensions and improvements on this work.

6.1 Research Conclusions

Identification of deceased individuals based on dental characteristics is receiving increased attention, especially with the large volume of victims encountered in mass disasters. In earlier work we proposed a prelude research prototype of an Automated Dental Identification System (ADIS), which we later extended and refined in collaboration with experts from the Criminal Justice Information Services (CJIS) Division of the US Federal Bureau of Investigation (FBI).

In this dissertation we addressed the following three problems pertaining to automating the dental identification process:

A) Automatic construction of dental charts:

Automatic classification of teeth into incisors, canines, premolars and molars, and hence automatic construction of dental charts, is essential for guiding the process of tooth-to-tooth matching. Dental charts capture the position of the radiographed teeth in the human mouth, thus they help avoiding illogical teeth comparisons that inefficiently consume the limited computational resources and may also mislead the decision-making process.

We presented a dual-stage approach for automatic construction of dental charts based on low computational-cost, appearance-based features and string matching. We initially independently classify each segmented tooth of a dental film based on teeth reconstruction in four image subspaces established using Principal Component Analysis (PCA). Then we validate and possibly correct initial teeth class assignments using teeth -neighborhood rules. Finally, if the resulting teeth class sequence is unique we assign a number for each tooth corresponding to its position in the dental quadrant it belongs to. Otherwise, if the resulting teeth class sequence is non-unique we call a reject option.

Testing results based on a sizable dataset suggest that our low-cost features achieve fast classification with an average film classification time of .8 seconds. Overall around 75% of the teeth are initially assigned to their proper classes, 86% of the teeth are assigned proper class labels after validation, and around 70% of the teeth assigned proper intra-quadrant numbers. The Accuracy of classification and numbering in bitewing films are approximately 5% higher than the overall results. By adjusting the penalization scheme for misclassification, we could achieve a teeth numbering accuracy of 82% in bitewing films with classification accuracy of 97% for premolars, and 89% for molars, also 93% of teeth were numbered within 1 neighbor with respect to the ground truth.

B) Micro and Macro Decision Making:

Dental radiographic records usually contain multiple films, which in turn show multiple views of teeth. To achieve reliable match decisions between a pair of cases based on dental characteristics, we devised a hierarchical fusion scheme that exploits two aspects of teeth multiplicity.

The lower level of our fusion scheme addresses the problem of combining multiple match scores of a tooth-pair due to the possible multiple views of the pair in the different radiographic films, and hence establishing a micro decision on the match status between a subject/reference tooth-pair.

The upper level of the hierarchical fusion scheme aggregates the micro decisions due to the different subject/reference teeth pairs to produce a macro decision on the match status between the subject record and the reference record. We also developed a scheme for ranking the different possible matches of a subject in a match list.

We instantiated three alternatives for micro decision-making based on the mean, median, and maximum rules for match score fusion. As macro decision-making relies on hardened micro decisions, we resorted to the majority-voting rule for case-to-case match decisions.

Testing results based on 397 case pairs, which involved comparison of 9900 teeth-pair views, show that the case-to-case genuine reject rate is around 15% based on the

mean and the max rules. The more conservative mean rule showed superiority in terms of higher hit rates near the top of the match lists, as well as a lower false accept rate.

C) Performance Limits of Dental Identification:

The performance of a biometric system depends, among other factors, on the capability of the feature representation scheme to capture the intra-subject invariance and the inter-subject variance, ideally retained by the biometric signal. We determined bounds on the capability of two sets of dental features, the first uses teeth contours, and the second is appearance-based.

We offered a generalized methodology that uses copulas to determine the number of degrees of freedom, as a figure of merit in performance analysis, which relaxes the stringent assumptions on feature independence when studying performance limits. We also presented workable approximations of the proposed approach.

6.2 Future Work

The research on achieving higher levels of automation in the dental identification process, beyond manual feature extraction and coding, is fairly recent and there are many interesting problems that have to be addressed in order to achieve the sought level of automation. In the following we list some of these problems.

- Automatic construction of dental charts: in some case films are assorted in a manner that does not immediately reflect the location teeth in the dental human atlas, for example, films may be flipped as well as translated either horizontally or vertically. To ensure proper construction of dental charts, detection of the maxilla and the mandible is essential. In practice, dentists rely on a physical impulse on films to determine the front face of the film and accordingly determine whether teeth are right side or left side teeth. Detection of this impulse may be a feasible problem from an image processing point of view. The remaining problem would be to determine whether the impulse surface touching the scanner was convex or concave. Along the line of using appearance based features for teeth classification, we think that Multiple

Discriminant Analysis (MDA) for the construction of *FisherTeeth* is an interesting approach that is worth pursuing.

- Fillings and partial teeth: the relatively high likelihood of physical transformation is a characteristic of dental features. Restorative fillings and partial fracture of teeth present two possibilities of physical transformations that manifest as change in the shape and content of teeth projections. Image inpainting⁴ is a candidate approach to work with teeth that were transformed by either one of the mentioned transformations.
- High-level features for efficient potential match search: so far all studied features in the context of the ADIS research prototype do not possess invariance properties that are needed for efficient retrieval. All matching approaches presented so far rely on alignment, which is a time consuming step. Invariant shape representations (e.g. moments, and also the alternative representations we described in section 5.4) present a possibility for achieving faster retrieval from a large database. However, we also have to bear in mind the possible physical changes in shape as pointed out earlier. Also detection of fillings is essential for ruling out potential mismatches due to impossibility of evolution from an AM to a PM situation (e.g. a filled AM molar may not evolve to a virgin PM molar). Moreover, detection of fillings and description of the restored surfaces, and hence automatic annotation of teeth with dental codes, present a significant improvement opportunity that uplifts the automation level in existing systems like CAPMI and WinID[®].
- Matching strategies: as we mentioned, films might be flipped vertically or horizontally, and unless these transformations are detected we cannot be very certain about the proper orientation of the dental chart, i.e. whether U_1 in Figure 4 corresponds to the upper right quadrant. One way around surviving with this situation is form four hypothesis regarding the proper orientation of a subjects dental chart,

⁴

We attribute this view for the problem to Dr. Xin Li.

perform matching (not necessarily exhaustive matching) under each of these hypotheses. Then select the strongest hypothesis for a more comprehensive case-to-case matching. We also recall that in some cases the number of tooth-pair views that we compared was unreasonable high, and in order not compromise on timeliness we may limit the number of views to be compared per tooth-pair.

- Error detection capabilities: each of the steps involved in processing of dental records is prone to error that likely to affect the overall performance of the system. In [8] aspect ratio and relative areas were used to exclude cropped regions that have apparent violations to expected values of these properties. In Chapter 6, we mentioned teeth classification ties as another form of error detection and as basis for exclusion of possible sources for performance degradation. We think that many segmentation errors may as well be detected by identifying a set of geometric and textural features possessed by true teeth. Also in alignment, we may use the resulting overlap area as basis to accept or reject registration.
- Performance limits based on matching multiple teeth: as we discussed in section 5.3.1, the intra-subject variability observed encountered in teeth appearance features reflect atypical characteristics of a biometric. However, teeth multiplicity is also a distinctive characteristic of dental records, which we argue to raise the upper limits on performance of dental identification based on appearance features. It would be interesting to conduct an empirical study to validate our argument. Such a study requires a large database of genuine matched teeth.
- 3D teeth views: so far dental identification using radiographs has been tackled using the 2D information represented in radiographic films. Because teeth are 3D objects, one may wonder whether availability of 3D image views would assist automation of the dental identification process. A complementary question to the former question is whether the multiple 2D views of a tooth, which are usually taken from slightly different angle from one another, would be sufficient to reconstruct a tooth in 3D.

In addition to the listed problems, we think that advocating imaging and digitizing standards in dental imaging will certainly aid automation the dental identification process.

Appendix: Tooth-to-Tooth Matching

We treat the problem of tooth-to-tooth matching as a binary classification problem, with the two classes being the Matched (M) class and the Unmatched class (\bar{M}). The classifier determines the probability of match between a pair of pre-registered regions of interest. This probability of match is computed as the posterior probability of match given evidence differences between the features extracted from an input ROI-pair; i.e. $p(M|\{|\Delta Z|\})$. The architecture of the tooth-to-tooth matcher resembles the pyramidal architecture shown in Figure 37. The components of this matcher are captured in the block diagram shown in Figure 38. This stage is composed of two layers (*see* Figure 37): the feature extraction layer and the Bayesian classification layer. The first layer comprises a bank of n_f 2D feature filters. Given an input ROI-pair (say ROI_A and ROI_B); the k^{th} filter maps each ROI to its corresponding 2D image in the k^{th} feature space thus, $ROI_A \xrightarrow{f_k} Z_A^{[k]}$, and $ROI_B \xrightarrow{f_k} Z_B^{[k]}$. In section A1 we elaborate on the feature extraction layer. The Bayesian classification layer computes the posterior probability of match between ROI_A and ROI_B given the difference evidence $\{|\Delta Z_{<x,y>}|^{[k]}, k = 1 \dots n_f\}$ between the features of ROI_A and ROI_B ($|\Delta Z_{(x,y)}|^{[k]} = |Z_{A(x,y)}^{[k]} - Z_{B(x,y)}^{[k]}|$ and $<x, y>$ denotes the entire k^{th} feature space). In order to compute the posterior probability of match, the class-conditional probability densities of the feature differences ($|\Delta Z|$) for both the Matched class and the Unmatched class $p_k(|\Delta Z|)$ and $q_k(|\Delta Z|)$ respectively have to be determined for all $k = 1 \dots n_f$. We elaborate on the Bayesian classification layer in section A2.

As in [75] and [76], we also take a supervised training approach in determining the parameters of the feature extraction filters and the parameters of class-conditional densities of the Bayesian classifier. However, we employ a novel training algorithm that combines back propagation with strategic searching of parameters using dynamic-range uni-variation of variables. We also extend the probability models of the class-conditional densities proposed by Baldi and Chauvin [75] in order to accommodate the nature of image variations within the dental regions of interest that we compare.

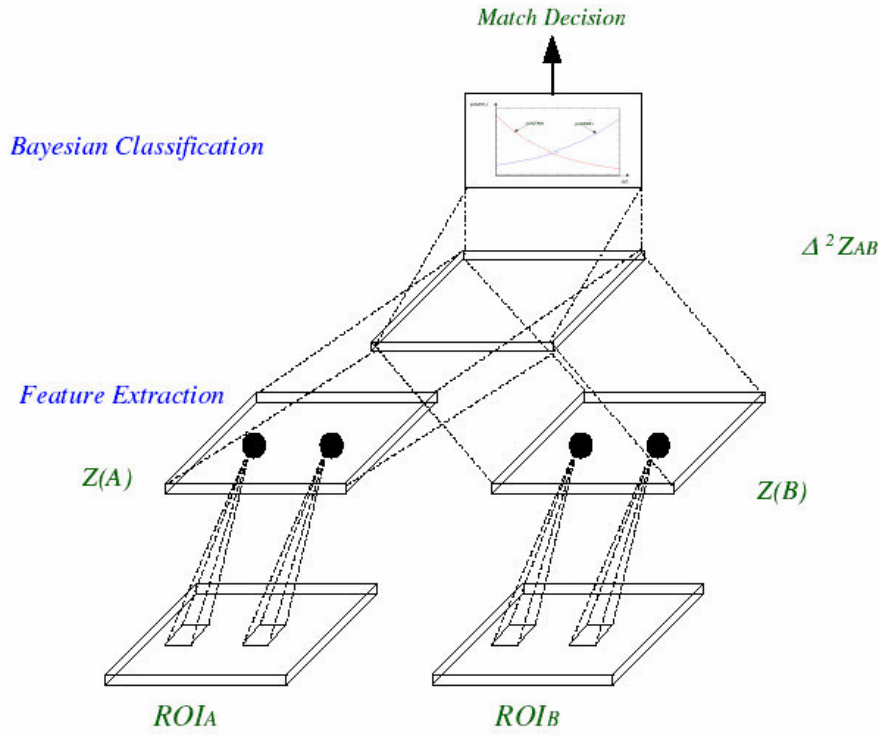


Figure 37: The pyramidal architecture for image matching.

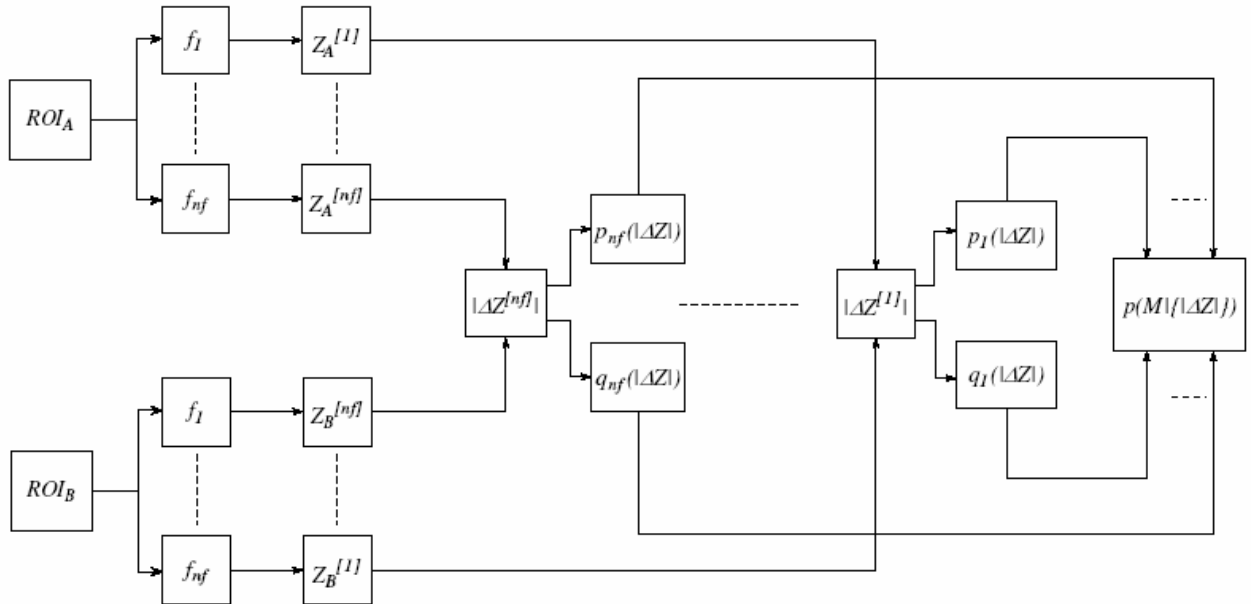


Figure 38: Block diagram of the feature extraction and classification stages.

A1 The Feature Extraction Layer

The feature extraction layer employs a set of nonlinear filters $\{f_k: k = 1, 2, \dots, n_f\}$ to map an $n \times n$ pixel ROI to a set of $m \times m$ pixel images in the corresponding feature spaces $\{Z^{[k]}: k = 1, 2, \dots, n_f\}$. Mapping of a given ROI to its image in the k^{th} feature space ($Z^{[k]}$) is achieved by the following nonlinear filtering operation:

$$Z_{(x,y)}^{[k]} = f(t_k, \xi_{(x,y)}^{[k]}) = \frac{1}{1 + e^{-(t_k - \xi_{(x,y)}^{[k]})}},$$

where:

$$\xi_{(x,y)}^{[k]} = \sum_{\langle x', y' \rangle} W_{(x', y')}^{[k]} I_{(x' + d(x-1), y' + d(y-1))}; \quad x = 1, 2, \dots, m, \quad y = 1, 2, \dots, m$$

$\xi_{(x,y)}^{[k]}$ is the discrete convolution (at (x, y)) of the k^{th} filter with the normalized and compressed ROI (I). By discrete convolution, here, we mean that the kernel does not slide over the entire ROI; rather it “hops” to designated locations on the ROI. Successive hop locations are d pixels apart from one another.

The parameters $W^{[k]}$, t_k of the k^{th} feature filter are the $(w \times w)$ coefficients of the convolution kernel, and the bias respectively. In each of the n_f spaces, the pixel values of feature images fall in the range $(0,1)$.

A feature image ($Z^{[k]}$) can be thought of as the output layer of a grid of $m \times m$ artificial neurons, each of which has a receptive field of $w \times w$ pixels. The receptive fields of neurons have some overlaps with those of neighboring neurons. These neurons share the weight set $W^{[k]}$, the bias t_k , and the binary sigmoid activation function f . This arrangement can also be thought of as a single neuron whose receptive field changes to cover the entire $n \times n$ normalized and compressed ROI.

As we previously mentioned, parameters of the filters evolve by learning during the training phase of the system. Thus, we do not explicitly specify the features to be used for matching; rather we present to the system a set of exemplar image ROI pairs, both matched and unmatched (or positive and negative examples). We adapt the filter parameters, and consequently change the features, so that the difference between features is reasonably small for matched exemplar pairs and the difference between features is reasonably large for

unmatched exemplar pairs. In section A4 we elaborate on the training algorithm we employ for estimation of parameters.

A2 The Bayesian Classification Layer

The Bayesian classification layer computes the posterior probability of match ($p(M | \{\Delta Z\})$) between a pair of ROI's using the differences between spatially corresponding features of the ROI-pair in all n_f feature spaces ($\{\Delta Z\} = \{\Delta Z_{<x,y>}\}^{[k]}$, $k = 1 \dots n_f$). To simplify notation, from this point on we use $\underline{\Delta}^{[k]}$ to denote $\{\Delta Z_{<x,y>}\}^{[k]}$, and $\underline{\Delta}$ to denote $\{\underline{\Delta}^{[k]}, k = 1 \dots n_f\}$. Using Bayes inversion formula, we may write $p(M | \underline{\Delta})$ as:

$$p(M | \underline{\Delta}) = \frac{p(\underline{\Delta} | M) \cdot p(M)}{p(\underline{\Delta})}$$

where $p(\underline{\Delta} | M)$ is the Matched class-conditional density of $\underline{\Delta}$, $p(M)$ is the prior probability of match, and $p(\underline{\Delta})$ is the evidence. Similarly, $p(\underline{\Delta} | \overline{M})$ denotes the Unmatched class-conditional density of $\underline{\Delta}$, and $p(\overline{M})$ denotes the prior probability of the unmatched class. Since the problem in hand is a binary classification problem then $p(\overline{M}) = 1 - p(M)$. By designating $p(\underline{\Delta} | M)$ as $\Phi(\underline{\Delta})$ and $\Theta(\underline{\Delta})$ as $p(\underline{\Delta} | \overline{M})$, the evidence $p(\underline{\Delta})$ can be expressed as:

$$p(\underline{\Delta}) = \Phi(\underline{\Delta}) \cdot p(M) + \Theta(\underline{\Delta}) \cdot p(\overline{M})$$

We assume that feature differences are independent random variables both across different feature spaces and within each feature space. Hence, we may write $\Phi(\underline{\Delta})$ and $\Theta(\underline{\Delta})$ as:

$$\begin{aligned} \Phi(\underline{\Delta}) &= \prod_{<k>} \Phi(\underline{\Delta}^{[k]}) = \prod_{<(x,y),k>} \Phi(\Delta_{(x,y)}^{[k]}) \\ \Theta(\underline{\Delta}) &= \prod_{<k>} \Theta(\underline{\Delta}^{[k]}) = \prod_{<(x,y),k>} \Theta(\Delta_{(x,y)}^{[k]}) \end{aligned}$$

Now we want to choose parametric models for the probability densities $\Phi(\Delta_{(x,y)}^{[k]})$ and $\Theta(\Delta_{(x,y)}^{[k]})$. In doing so, we seek $\Phi(\Delta_{(x,y)}^{[k]})$ -the Matched class-conditional density- to be a decreasing function of $\Delta_{(x,y)}^{[k]}$ and $\Theta(\Delta_{(x,y)}^{[k]})$ -the Unmatched class-conditional density- to be

an increasing function of $\Delta_{(x,y)}^{[k]}$. Similar to the models proposed in [75], we propose distributions of a binomial form. However, we augment these models by introducing parameters for scaling of the feature differences. The rational for introducing these scaling parameters is that the computed evidence, $\Delta_{(x,y)}^{[k]}$, need not be treated evenly by $\Phi(\Delta_{(x,y)}^{[k]})$ and $\Theta(\Delta_{(x,y)}^{[k]})$. Based on examinations of several dental radiographs, we observed that unmatched ROI-pairs may have areas over which the evidence, $\Delta_{(x,y)}^{[k]}$, is fairly small (For example, $\Delta_{(x,y)}^{[k]}$ is significantly small for some spatially corresponding bright areas within the boundaries of unmatched teeth). Without these evidence-scaling parameters, we found that convergence of training (estimation of system parameters) was impossible.

The models we propose for $\Phi(\Delta_{(x,y)}^{[k]})$ and $\Theta(\Delta_{(x,y)}^{[k]})$ are:

$$\begin{aligned}\Phi(\Delta_{(x,y)}^{[k]}) &= \phi_k^{1 - \alpha_k \Delta_{(x,y)}^{[k]}} \cdot (1 - \phi_k)^{\alpha_k \Delta_{(x,y)}^{[k]}} \\ &= \phi_k \cdot \left(\frac{\phi_k}{1 - \phi_k} \right)^{-\alpha_k \cdot \Delta_{(x,y)}^{[k]}} \\ \Theta(\Delta_{(x,y)}^{[k]}) &= \theta_k^{\beta_k \Delta_{(x,y)}^{[k]}} \cdot (1 - \theta_k)^{1 - \beta_k \Delta_{(x,y)}^{[k]}} \\ &= (1 - \theta_k) \cdot \left(\frac{\theta_k}{1 - \theta_k} \right)^{\beta_k \cdot \Delta_{(x,y)}^{[k]}}\end{aligned}$$

where ϕ_k , θ_k , α_k , β_k are classifier parameters that we estimate by supervised training. The scaling parameters we introduced are α_k , β_k .

A3 Decision Making

The match decision is made based on the computed value of the posterior probabilities $p(M | \underline{\Delta})$ and $p(\overline{M} | \underline{\Delta})$, according to the following likelihood ratio rule:

$$Decide : \begin{cases} C_i & \text{if } \frac{P(C_i | \underline{\Delta})}{P(C_{1-i} | \underline{\Delta})} \geq \rho; i \in \{0,1\} \\ Undetermined & \text{otherwise} \end{cases}$$

where C_0 designates the Unmatched class (\overline{M}), C_1 designates the Matched class (M), and ρ is the minimum likelihood ratio, below which we think an accurate decision can not be made and call out the “Undetermined” -or rejection- option. In choosing the value of ρ , we seek to achieve maximum separation between the two classes, while minimizing the false rejection rate (FRR).

A4 The Training Algorithm

We have earlier presented the structure of the tooth-to-tooth matcher and the mathematical models we propose for this component of the image comparison matching. In this section we introduce a hybrid training-algorithm for determining the parameters of the feature filters and the class-conditional densities of the binary Bayesian classifier.

A4.1 Training Objective

During the training phase, we use exemplars for estimating values of the matcher parameters. The estimated parameter values should achieve certain classification accuracy requirements. Generally, training of a neural network involves an optimization problem in which a cost function is to be minimized according to some constraints. In our case, the output of the matcher is to be interpreted as a probability, and a suitable cost function would be the cross entropy between the estimated probability, $p = p(M|\Delta Z)$, and the true probability assigned to an exemplar pair, P , summed over the set of all training exemplars. Note that $P = 1$ for matched exemplar pairs and $P = 0$ for unmatched exemplar pairs. The cross entropy function is one of the information theoretic measure of discrepancy between probability distributions [78][79], and is given by:

$$H(p, q) = \sum_{\langle A, B \rangle} (P \log \frac{P}{p} + Q \log \frac{Q}{q})$$

Where $q = 1 - p$, $Q = 1 - P$. H is a non-negative convex function of p and q , and H equals to zero if and only if $p = P$ and $q = Q$ for all training pairs.

A4.2 Back Propagation

The adjustment of neural network weights using back propagation of observed error at the network output is the most famous technique for supervised training of neural networks. The idea of back propagation is to use gradient descent in optimizing the cost function. An important factor in achieving convergence in any gradient descent technique is proper learning rates for the different weights. Moreover, for a given weight, it might be necessary to dynamically change the learning rate during the learning process so that convergence would be more probable. A too high learning rate leads to missing minima and causes overshoots that will transcend the neighborhood of the well of attraction around a local or a global minimum [41][80]. On the other hand, a too low learning rate results in too slow convergence and may lead to trapping the learning process in an unsatisfactory local minimum [41][81]. Several techniques for choosing variable and dynamic learning rates while training by gradient descent are presented in [41][80][81], however, there are no guarantees on convergence. Incorporation of a momentum term tends to provide a smoother descent toward a minimum by suppressing excessive oscillations in the cost function values starting from a given point heading toward a minimum point. We utilize back propagation as one of two learning techniques incorporated in the hybrid-training algorithm that we present in A4.4. Different training modes are possible as described in [41][77]; we chose the pattern mode strategy, in which parameters are updated after presentation of each exemplar pair.

A4.3 Adaptive Uni-Variation of Weights

In order to improve the convergence rate of the training phase, we introduce an adaptive strategic searching technique for weight adjustment. A non-adaptive version of the technique that we propose in this section is described in [80] as a strategic searching approach for weight initialization in neural networks. Our contribution provides search over a dynamic range for each weight, which extends the search space and make the technique valuable beyond weight initialization.

We treat each weight individually as implied by “uni-variation” of weights. As any other training method, the target of our proposed searching technique is to find a weight set that minimizes a convex cost function. We suggest that the cost function (H) assumes quadratic

dependence on each weight individually. Hence, for each weight (ψ) we search for the point (ψ_{\min}) that minimizes the function $H_2(\psi)$, which takes the form $H_2(\psi) = a\psi^2 + b\psi + c$. $H_2(\psi)$ is the assumed quadratic approximation of $H(\psi)$ when the dependence of H on any weight other than ψ is neglected.

With all weights other than ψ kept unchanged, we evaluate the cost function H at three different values of ψ ; namely ψ_{-1} , ψ_0 , and ψ_1 . We then fit the obtained points $\{(\psi_i, H(\psi_i)): i = -1, 0, 1\}$ to the curve $H_2(\psi)$. We chose ψ_0 to be the current value for the weight ψ , thus saving the time of one function evaluation for each weight. We also choose $\psi_{-1} = 0.9 \psi_0$ and $\psi_1 = 1.1 \psi_0$ provided that $\psi_0 \neq 0$. This choice of values allow for an adaptive search unlike the search over static range described in [80]. To determine the point ψ_{\min} , we consider two cases based on the value of the coefficient 'a': (i) when $a > 0$, $\psi_{\min} = -b/2a$, (ii) when $a \leq 0$, ψ_{\min} is either ψ_{-1} or ψ_1 depending on the values of H_{-1} and H_1 .

After determining the value of ψ_{\min} , it replaces the current value of ψ in the weight set. Then, the cost function H is evaluated, if a decrease in its value is encountered, the new value of the weight ψ is affixed, otherwise the its old value is restored. Figure 39 illustrates the hybrid training technique in algorithmic steps.

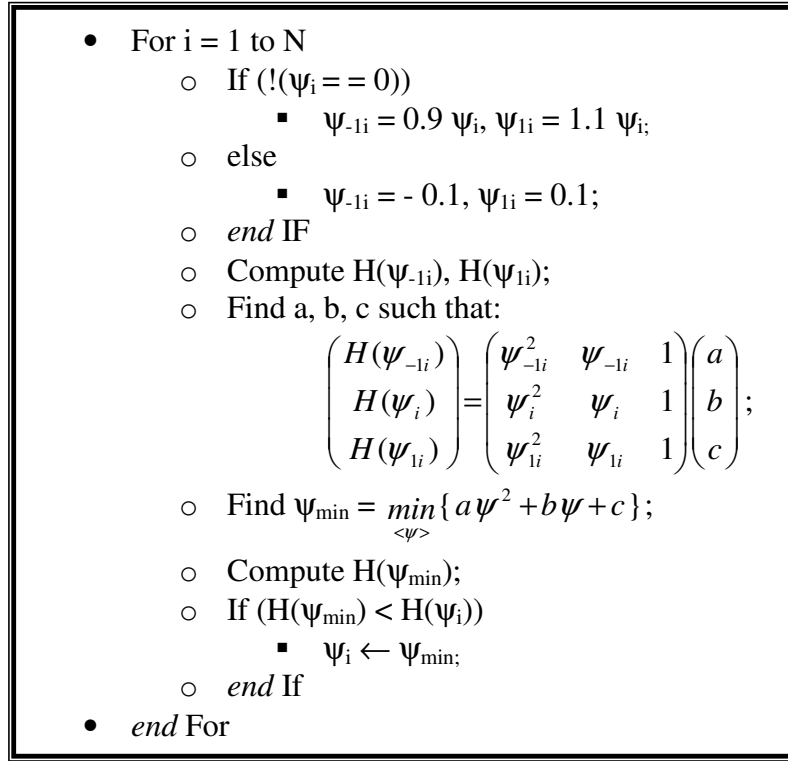


Figure 39: The adaptive uni-variation of weights in pseudo code

This search process is carried out for each weight (or parameter) of the feature extraction and classification layers. Obviously, a uni-variation approach is computationally more expensive than back propagation; as the number of cost function evaluations per training iteration in the former is three times the number of the parameters to be estimated, while in batch mode back propagation [41] the cost function is evaluated only once per training iteration.

Accordingly, we resort to this technique only when we observe that back propagation seems to be trapped in an unsatisfactory solution. The hybrid training technique outlined in section A4.4 switches between back propagation and adaptive uni-variation of weights based on observation of training progress.

A4.4 A Hybrid Training Approach

The training process continues as long as the stopping criterion is not met, provided that there is progress towards a satisfactory minimum of the cross entropy function. A satisfactory minimum is associated with a maximum permissible error in the estimated posterior matching probability over the entire set of exemplars.

In the results section, we provide give our observations on the role of back propagation and adaptive uni-variation of weights in achieving the training objective. Figure 40 illustrates the hybrid training technique in algorithmic steps. Priority is given to training by back propagation because it is computationally less expensive than adaptive uni-variation. During the training process, when back propagation is found not to achieve progress towards the training target, our strategic searching technique overrides. The strategic searching technique works on moving the weight set towards a more suitable point in the weight space before back propagation is resumed.

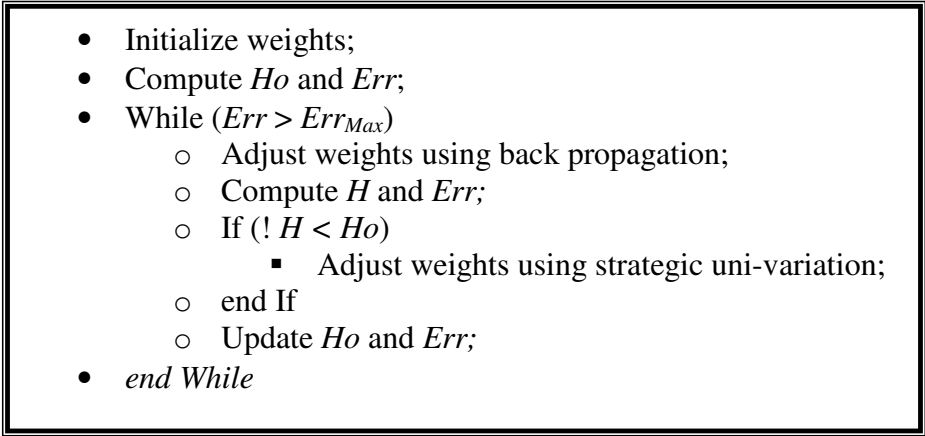
- 
- Initialize weights;
 - Compute H_o and Err ;
 - While ($Err > Err_{Max}$)
 - Adjust weights using back propagation;
 - Compute H and Err ;
 - If ($! H < H_o$)
 - Adjust weights using strategic uni-variation;
 - end If
 - Update H_o and Err ;
 - end While

Figure 40: The hybrid-training algorithm in pseudo code.

A5 Experimental Results

In this section we report the results of training and testing experiments we carried out to instantiate and verify the tooth-to-tooth matching stage of the image comparison component of the prototype ADIS. The results we present in this section are all based on un-compiled MATLAB[®] realizations running on a 256 MB RAM, 1.13 GHz Intel P3[®] based personal computer. The dental radiographs used for both training and testing were extracted from the digitized dental images database provided by the CJIS [64][65]. We chose to restrain our training and testing to periapical and bitewing films [82], which usually capture the distinctive dental features at reasonable level of detail.

A5.1 Non-learnable Parameters

As we mentioned in section A1 the feature extraction layer is composed of a filter bank of n_f feature filters, which map an $n \times n$ pixel ROI to n_f feature images each of which is $m \times m$ pixels. Each $m \times m$ feature image is produced by a nonlinear operation on the output of discrete convolution of a $w \times w$ feature-kernel with the input ROI. We chose: (i) the number of feature filters to be $n_f = 2$, (ii) compressed ROI size of 32×32 pixels, (iii) the size of a feature image to be 6×6 , and (iv) the feature-kernel size to be 7×7 . We made these choices in based on rationale and experiments; we want to limit the number of learnable parameters in order not to experience the *curse of dimensionality* [41], meanwhile we do not want to sacrifice the discrimination power of the features and hence the accuracy. In section A3, we proposed that decision-making be based on comparison of the maximum likelihood ratio with a desirable ratio (ρ), in our experiments we have chosen $\rho = 4$. Based on several training trials under different choices of these non-learnable parameters, we found that this combination gives good results.

A5.2 Training

Given the values of the non-learnable system parameters that we mentioned in section A5.1, the number of learnable parameters is 108 as follows: 49 kernel weights and 1 sigmoid bias for each filter, and 4 parameters for the Bayesian classifier associated with each feature filter.

We conducted several training experiments using training sets with different number of training pairs and for which we varied the ratio of matched pairs to the number of training pairs. In choosing exemplars we made sure that ROIs capture teeth root areas, which are believed to be stable features. Additionally, we have included exemplars that originated from non high-quality images. Figure 41 shows examples of pairs used for training; Figure 41 (a) shows example of an unmatched pair, and Figure 41 (b) shows example of a matched pair. Note that we have intentionally introduced horizontal and vertical translations between the matched pair in Figure 41 (b); that is we provided some noise in the training set in order to achieve some robustness.

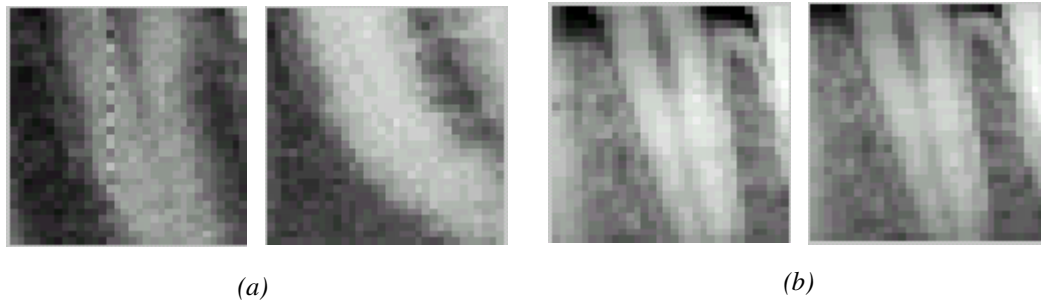


Figure 41: Some training pairs (a) unmatched pair, (b) matched pair.

A5.2.a Convergence

As we mentioned in section A2, *Baldi* and *Chauvin* [75] suggested binomial distributions for the class-conditional probability density functions (pdfs) of the decision layer of their fingerprint identification system. These binomial pdfs imply that the class-conditional distributions $p(\underline{\Delta}|M)$ and $p(\underline{\Delta}|\overline{M})$ are allowed a slight difference in how they *respond* to an observed feature difference. After several extended-period training attempts, which involved variations of training techniques and parameters, we observed that the separation between the *Matched* and *Unmatched* classes was *too narrow*. However, it was clear that there was a fine line that separated the two classes. This narrow separation is mainly due to slight feature-differences for some spatially corresponding areas within the boundaries of unmatched teeth. Therefore, we introduced the feature-difference scaling parameters α and β in the models of the class-conditional densities.

With the scaling parameters (α and β), the training of the matcher converged to a solution with remarkably low classification error. Column 5 of Table 9 shows that the average classification error is in an order of magnitude less than or equal 10^{-7} . We also observed that convergence was reached mainly by virtue of our adaptive technique for uni-variation of weights described in A4.3. Figure 42 shows the convergence profile of the matcher based on the second training set described in Table 9, the -markers indicate back propagation cycles of the hybrid training algorithm we proposed in section A4.4, while the *-markers indicate the adaptive uni-variation counterpart of the hybrid algorithm. As can be seen from Figure 42, back propagation did significantly contribute in steering the cross entropy function towards a satisfactory minimum. However, we believe that allowing more dynamics in

learning by back propagation may improve its share in achieving convergence (e.g. use of variable learning rates).

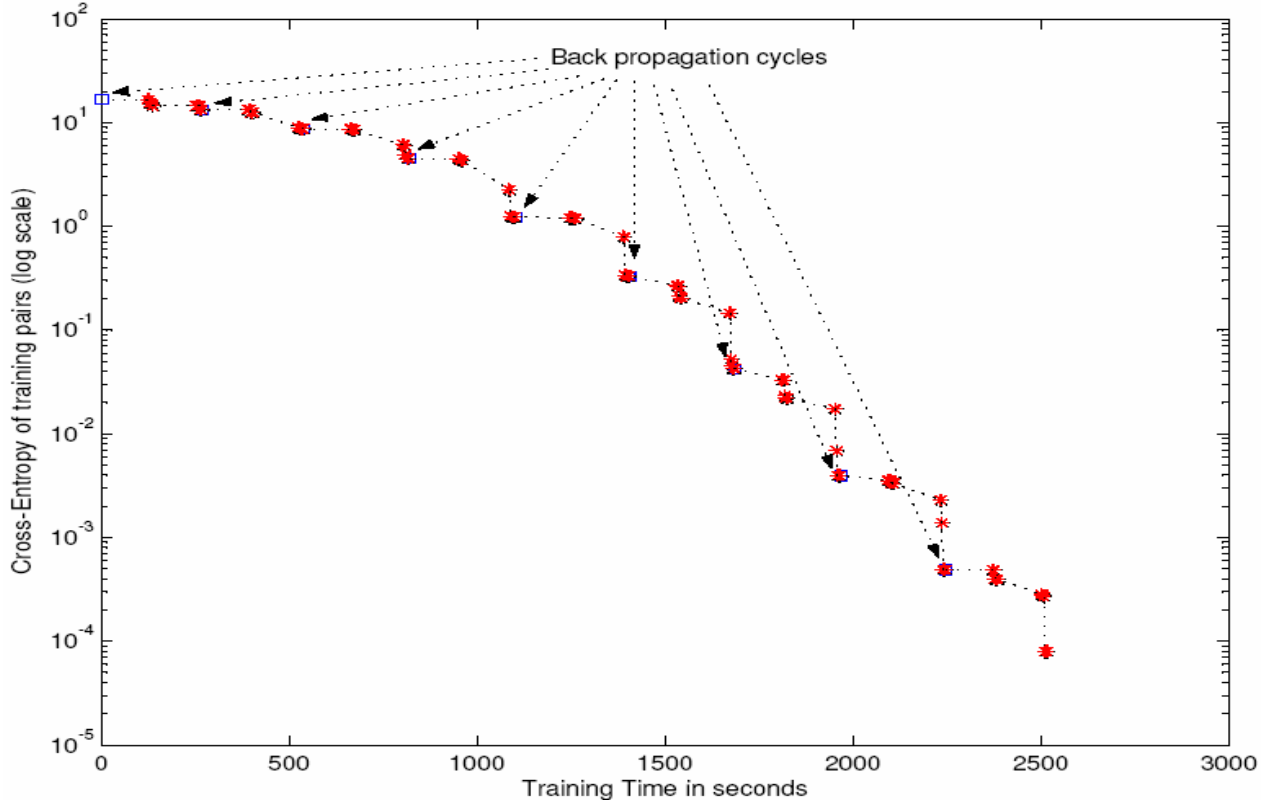


Figure 42: Convergence profile of the matcher using the hybrid algorithm in section A4.4

A5.2.b Examples of Training Experiments

Table 9 illustrates two examples of training experiments. N_{train} is the number of training pairs in the experiment, H_{min} is value of the cross entropy function reached at the end of training, $\text{error}_{\text{avrg}}$ is the average error in estimated matching probabilities for the entire training set after convergence, and $T_{\text{convergence}}$ is the training time in minutes. It is worth noting that number of training pairs N_{train} is actually the exemplar set used for the Bayesian classifier, however, the number of exemplars used in training the feature extraction layer is $2m^2 N_{\text{train}}$ as each ROI incorporates m^2 receptive fields.

The stopping criterion we used in both experiments was achieving a maximum permissible of 1% over the entire respective training set. For a given training pair, we compute the training error as: $\text{error} = |P - p|$, where P is the match *state-of-nature* between the training pair ($P = 1$

for a matched pair, and $P = 0$ otherwise), and p is the estimated posterior probability of match.

While the maximum permissible error was set to 1%, the achieved accuracy was always better than the specified level by at least two orders of magnitude, which means that the maximum training error was always less than 0.01% [6]. The figures given in Table 9 are typical of those we obtained in several training experiments.

Table 9: Examples of training experiments for the tooth-to-tooth matcher.

Training Set No.	N_{train}	% Match	$H_{\min} \times 10^{-8}$	$\text{error}_{\text{avg}} \times 10^{-9}$	$T_{\text{convergence}}$ (sec.)
1	15	33.3	10.131	6.7538	≈ 400
2	30	20	1.7×10^3	5.7×10^2	≈ 2500

A5.2.c An Interpretation of Feature Extraction Filters

Figure 43 shows a ROI and image representations of its extracted features ($Z_{\text{ROI}}^{[1]}$, $Z_{\text{ROI}}^{[2]}$). By visually inspecting $Z_{\text{ROI}}^{[1]}$ and $Z_{\text{ROI}}^{[2]}$ we may infer that the first feature extraction filter produces a coarse gradient image of the ROI, while the second filter produces coarse inversion of the ROI. We also note that the shape of the root in Figure 43(a) is captured in both feature images.

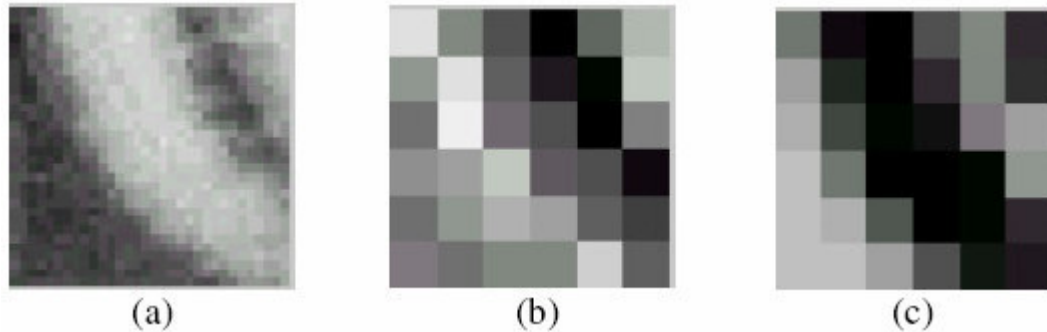


Figure 43: Example of a ROI (a) and its extracted feature images; $Z_{\text{ROI}}^{[1]}$ (b) and $Z_{\text{ROI}}^{[2]}$ (c).

A5.3 Testing

In this section we present results of testing the trained matcher. We consider two venues in assessing the performance of this stage: (i) testing with a genuine dataset of ROI pairs manually preprocessed from AM and PM radiographs whose match states-of-nature are known, and (ii) testing with synthetic datasets of ROI pairs whose match states-of-natures are

determined based on the preprocessing steps that we perform on the dental radiographs in order to generate these ROI pairs. While the first venue is the ultimate on assessing the performance of this stage, however, lack of a considerable number of cases with ground truth matches motivates us to also resort to the second type of testing in order to obtain statistically sound results. Studying the sensitivity of the matcher to imperfections in ROI-pair pre-registration adds another motivation to testing with synthetic datasets.

A5.3.a Testing with a Genuine Dataset

The dataset we used in this type of testing is extracted from known matches between AM and PM records on CJIS database [65], the total number of these cases is only 14. Out of these 14 cases we segmented 51 ROI pairs and used them as the base of the genuine dataset. We annotated each ROI with a label corresponding to location of the tooth it corresponds to. We used this label to identify candidate ROIs to compare a subject ROI with; for example if a subject ROI is a crown of a PM tooth in the right segment of the maxillary plane (upper jaw), then we choose its candidates to be AM teeth in the right segment of the maxillary plane. The total number of resulting ROI pairs is 189 out of which 51 pairs are genuine matches.

Table 10 summarizes the outcome of testing with the genuine dataset described above. Note that the false reject rate (FRR) in this case does not reflect the actual FRR of the system, as the number of genuinely matched ROI-pairs is small. As a matter of fact genuine testing with 189 pairs does provide sound statistics.

Table 10: Results of testing the matcher using a genuine dataset

Accuracy	Error			Undetermined
	FRR	FAR	Total	
82 %	0 %	8 %	8 %	10 %

A5.3.b Testing with Synthetic Datasets

The datasets we used in this type of testing are extracted from missing (AM) and unidentified (PM) cases on CJIS databases [64][65], which contain about two thousand cases that are believed to have no matches. In generating datasets for this type of testing we: (i) randomly choose a ROI from a dental radiograph, and (ii) choose another ROI either from the same dental radiograph or from a different one as we describe hereafter. We choose the ROI size to

be 128×128 pixels, a size which we found to sufficiently contain distinctive features in all images that we used.

To generate matched test ROI-pairs we choose the both ROIs from the same dental radiograph such that the Euclidean distance between their centers is within 10 pixels. On the other hand, we generate unmatched ROI-pairs by choosing them either from different dental radiographs, or from the same dental radiograph such that the Euclidean distance between their centers is at least 20 pixels.

We generate thousands of test ROI-pairs using the method specified above. We present test datasets -after ROI compression and normalization- to the matcher and compare the obtained decisions to the Matched/Unmatched labels that we assigned to test ROI-pairs of the dataset.

To study the sensitivity of the toot-to-tooth matcher to possible *slight* misalignments between ROI-pairs, we also test the matcher under scaling and rotation transformations of ROI-pairs of synthetic datasets.

In Table 11 we present a sample of typical testing results with synthetic datasets, which were formed by vertical and horizontal translations according to the Euclidean distance constraints we described above. The second and third columns of Table 11 give dataset characteristics; N_{test} is the total number of ROI-pairs used in the respective experiments, and match is the percentage of ROI-pairs labeled as matched amongst these N_{test} pairs. The forth column identifies the weight set we used in conducting each experiment, weight sets 1 and 2 correspond to those evolved from training sets 1 and 2, respectively in Table 9. The last five columns indicate the outcome of experiments in terms of accuracy, error rates, and percentage of undetermined decisions.

In experiments 1 and 2 we used the same dataset but with different weight set, accuracy is slightly better with the weight set 1, however, FRR is better with weight set 2 and hence we used it with the remainder of experiments presented in Table 11. In experiments 3 and 4 we varied the percentage of matched pairs in the test datasets in order to check for biases in decisions, based on performance measures we do not see significant biases. Experiments 5 and 7 present testing with extreme datasets in terms of percentage match amongst the test ROI-pairs. Experiment 6 is another example performance with a balanced dataset in terms of the percentage of matched test ROI-pairs. Overall we clearly see accuracy in excess of 90 %.

Table 11: Results of testing the matcher using synthetic datasets.

Experiment	Dataset C/Cs		Weight set	Accuracy	Error			Undetermined
	N_{test}	match			FRR	FAR	Total	
1	5000	50 %	1	98.22 %	1.20 %	0.04 %	1.24 %	0.54 %
2	5000	50 %	2	97.94 %	0.12 %	1.28 %	1.40 %	0.66 %
3	10000	10 %	2	98.68 %	0.13 %	0.53 %	0.66 %	0.66 %
4	10000	30 %	2	97.21 %	1.89 %	0.00 %	1.89 %	0.90 %
5	20000	00 %	2	99.89 %	N/A	0.07 %	0.07 %	0.04 %
6	20000	50 %	2	95.31 %	3.21 %	0.03 %	3.24 %	1.46 %
7	20000	100 %	2	90.75 %	6.53 %	N/A	6.53 %	2.72 %

Figure 44 summarizes a set of experiments in which 1000 ROI-pairs, 50% matches, were subject to intentional $\pm 15\%$ contrast in scaling, in steps of 5%. We observe that over a range of $\pm 5\%$ difference of scaling, accuracy is mostly unaffected. Meanwhile, we also observe that over a range of $\pm 10\%$ difference of scaling, the FRR is below 10 %.

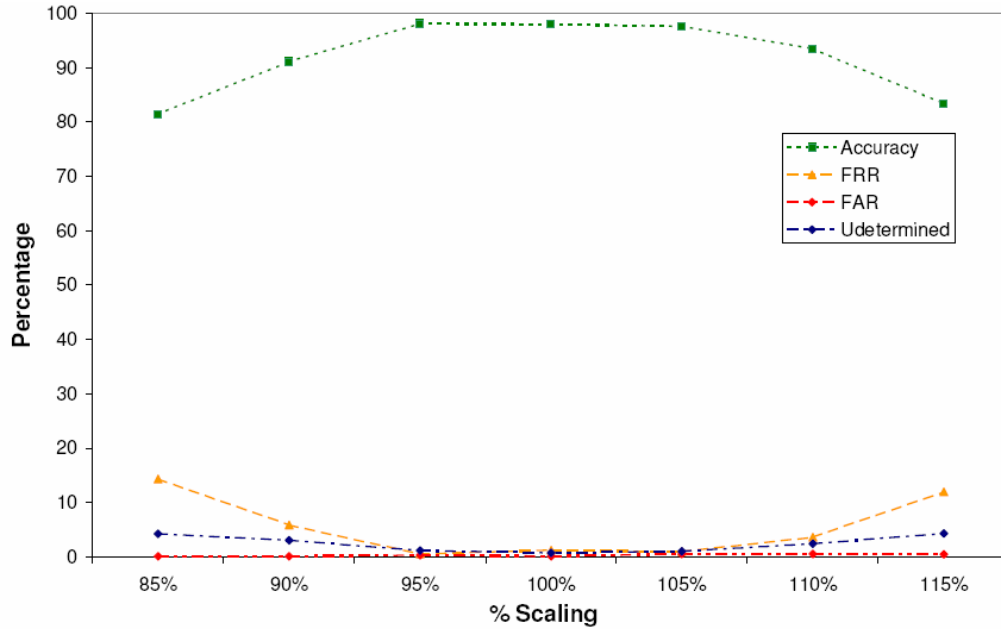


Figure 44: Performance of the tooth-to-tooth matcher under scaling variations.

Figure 45 summarizes a set of experiments in which 5000 ROI-pairs, 50% matches, were subject to intentional $\pm 15^\circ$ misalignments, in steps of 1° . We observe a severe degradation in accuracy for any rotational misalignment. However, over the range of $\pm 5^\circ$ this degradation is predominantly due to increased FAR while the FRR is mostly marginal. We also observe that that over a range of $\pm 10^\circ$ of rotational misalignments, the FRR is below 10 %.

We believe that this sensitivity analysis suggests that the matcher bears robustness against *slight* imperfections in the preprocessing stage. We recall that the exemplars we presented to this stage during the training phase did not incorporate any misalignment due to scaling or rotation, a fact that makes us further appreciate this achieved robustness.

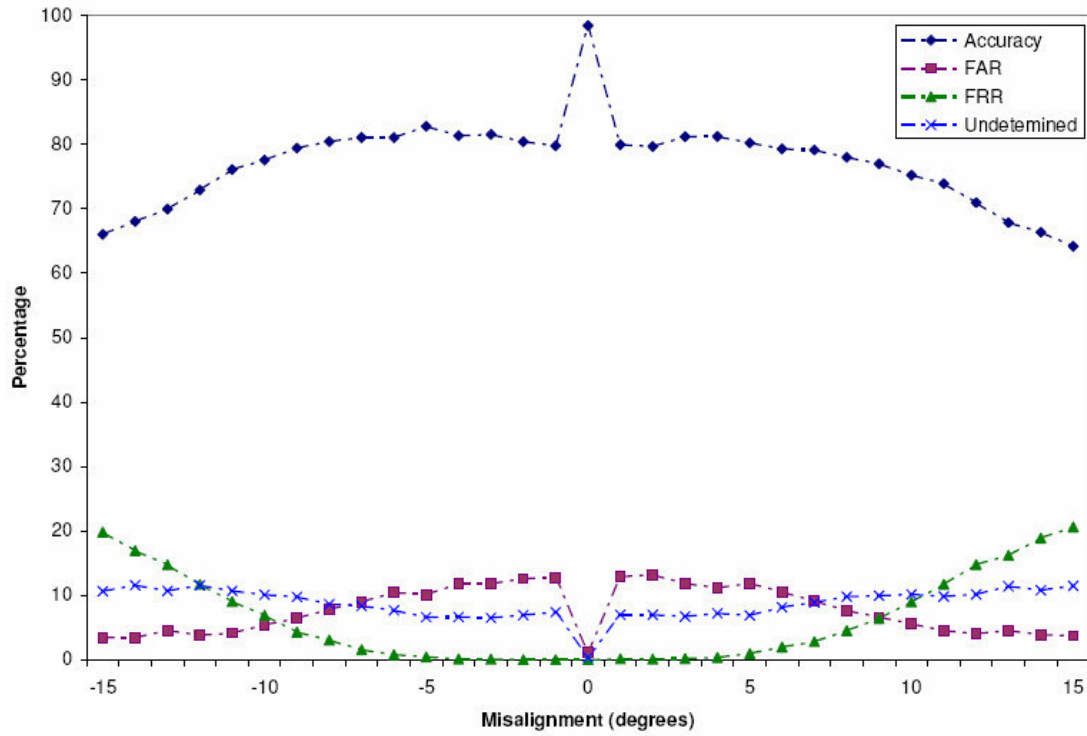


Figure 45: Performance of the tooth-to-tooth matcher under rotation variations.

Bibliography

- [1] P. Stimson & C. Mertz, *Forensic Dentistry*. CRC Press 1997
- [2] American Society of Forensic Odontology, *Forensic Odontology News*, vol. 16, no. 2, Summer 1997
- [3] D. F. MacLean, S. L. Kogon, and L. W. Stitt, "Validation of Dental Radiographs for Human Identification," *Journal of Forensic Sciences*, JFSCA, vol. 39, no. 5, Sep. 1994, pp. 1195-1200.
- [4] The Canadian Dental Association, *Communique*, May/June 1997.
- [5] United States Army Institute of Dental Research Walter Reed Army Medical Center, "Computer Assisted Post Mortem Identification via Dental and other Characteristics", *USAIDR Information Bulletin*, vol. 5, no. 1, Autumn 1990.
- [6] Diaa Eldin Nassar, "A Prototype Automatic Dental Identification System (ADIS)". *Masters Thesis*, Department of Electrical and Computer Engineering - WVU, Apr. 2001.
- [7] G. Fahmy, D. Nassar, E. Haj-Said, H. Chen, O. Nomir, J. Zhou, R. Howell, H. H. Ammar, M. Abdel-Mottaleb and A. K. Jain, "Towards an automated dental identification system (ADIS)", *Proc. ICBA (International Conference on Biometric Authentication)*, Hong Kong, Jul. 2004, pp. 789-796.
- [8] Xin Li, Ayman Abaza, Diaa Eldin Nassar, and Hany Ammar, "Fast and Accurate Segmentation of Dental X-ray Records", to appear in *Proc. ICBA 2006 (the 2006 International Conference on Biometric Authentication)*, Hong Kong, Jan. 2006.
- [9] Eyad Haj-Said, Diaa Eldin Nassar, Gamal Fahmy, and Hany Ammar, "Dental X-ray Image Segmentation" to appear in *IEEE Transactions on Information Forensics and Security*.
- [10] Jindan Zhou, and Mohamed Abdel-Mottaleb, "A Content-based System for Human Identification based on Bitewing Dental X-Ray Images", *Pattern Recognition*, 38(2005), pp. 2132-2142.
- [11] Anil K. Jain and Hong Chen, "Matching of Dental X-ray Images for Human Identification", *Pattern Recognition*, vol. 37, no. 7, Jul. 2004, pp. 1519-1532.

- [12] Hong Chen and Anil K. Jain, "Tooth Contour Extraction for Matching Dental Radiographs", *Proc. ICPR 2004*, vol. III, pp. 522-525, Aug. 2004, Cambridge, UK.
- [13] Omaima Nomair and Mohamed Abdel-Mottaleb, "A system for human identification from X-ray dental radiographs", *Pattern Recognition*, 38(2005), pp. 1295-1305.
- [14] Mohammad Mahoor and Mohamed Abdel-Mottaleb, "Automatic Classification of Teeth in Bitewing Images", *Proc. of the ICIP 2004*, Singapore Oct. 2004
- [15] Anil K. Jain, and Hong Chen, "Registration to Dental Atlas", *Proc. of the SPIE Biometric Technology for Human Identification II*, vol. 5779, , Mar. 2005, Orlando, Florida, pp. 292-298.
- [16] H. Chen and A.K. Jain, "Dental Biometrics: Alignment and Matching of Dental Radiographs", *IEEE Transactions on PAMI*, vol. 27, no. 8, Aug. 2005, pp. 1319-1326.
- [17] Zainab N. Millwala, "A Dual Stage Approach for Dental Image Registration", *Masters Thesis*, Lane Department of Computer Science and Electrical Engineering, West Virginia University, Dec. 2004.
- [18] Mythili Ogirala, "Multi Resolution Dental Image Registration based on Genetic Algorithm", *Masters Thesis*, Lane Department of Computer Science and Electrical Engineering, West Virginia University, May 2005.
- [19] Diao Eldin Nassar, Zainab Millwala, Donald Adjero, and Hany Ammar, "Dual-Stage Dental Radiograph Alinment", submitted to the *Journal of Forensic Sciences*.
- [20] Diao Eldin Nassar, Mythili Ogirala, Donald Adjero, and Hany Ammar, "An Efficient Multi-Resolution GA approach to Dental Image Alignment", to appear in *Proc. of the 2006 SPIE Electronic Imaging*.
- [21] D. Nassar, and H. Ammar, "A Neural Network based System for Dental Radiograph Comparison", *2nd IEEE International Symposium on Signal Processing and Information Technology (ISSPIT2002)*, Dec. 2002, Marrakech, Morocco.
- [22] Dr. Jim McGivney et al. WinID2[®] software. <http://www.winid.com>
- [23] L. Lorton, M. Rethman, and R. Friedman, "The Computer-Assisted Postmortem Identification (CAPMI) System: A Computer-Based Identification Program," *Journal of Forensic Sciences*, vol. 33, no. 4, July 1988, pp. 977-984.

- [24] A. B. Williams, R. B. Friedman, and L. Lorton, "A New Algorithm for Use in Computer Identification," *Journal of Forensic Sciences*, vol. 34, no. 3, May 1989, pp. 682-686.
- [25] L. Lorton, M. Rethman, and R. Friedman, "The Computer-Assisted Postmortem Identification (CAPMI) System: Sorting Algorithm Improvements," *Journal of Forensic Sciences*, vol. 34, no. 4, July 1989, pp. 996-1002.
- [26] R. B. Friedman, K. A. Cornwell, and L. Lorton, "Dental Characteristics of a Large Military Population Useful for Identification," *Journal of Forensic Sciences*, vol. 34, no. 6, Nov. 1989, pp. 1357-1364.
- [27] Jim McGivney, Commentary on: Cheri Lewis, "Win ID2 versus CAPMI4: two computer-assisted identification systems", *Journal of Forensic Sciences*, 2002; 47(3): 536-538. *Journal of Forensic Sciences*, 2003; 48(2): 472.
- [28] Cheri Lewis, and Les Leventhal, "Locator System Versus WinID3 Versus CAPMI4: Identifying Victims from Dental Remains in a Large Disaster", *Journal of Forensic Identification*, vol. 54, no. 2, 2004 1984, pp. 185-202.
- [29] C. M. Bowers, R. J. Johansen, "Digital imaging methods as an aid in dental identification of human remains", *Journal of Forensic Sciences*, 2002;47(2):354-359.
- [30] H. Soomer, M. Lincoln, H. Ranta, A. Penttilä, and E. Leibur, "Dentists' Qualifications Affect the Accuracy of Radiographic Identification," *Journal of Forensic Sciences*, vol. 48, no. 5, Sep. 2003, pp. 1121-1126
- [31] S. M. Las Heras, A. Valenzuela, C. Ogayar, A. J. Valverde, and J. C. Torres, "Computer-Based Production of Comparison Overlays from 3D-Scanned Dental Casts for Bite Mark Analysis", *Journal of Forensic Sciences*, 2005;50(1):127-133.
- [32] R. D. Rawson, R. K. Ommen, G. Kinard, J. Johnson, and A. Yfantis, "Statistical Evidence for the Individuality of the Human Dentation," *Journal of Forensic Sciences*, JFSCA, vol. 29, no. 1, Jan. 1984, pp. 245-253.
- [33] R. Veltkamp and M. Hagedoorn, "State-of-the-Art in Shape Matching", *Technical Report UU-CS-1999-27*, Utrecht University, Utrecht (The Netherlands) - Department of Computer Science 1999.
- [34] Lisa G. Brown, "A Survey of Image Registration Techniques". *ACM Computing Surveys*, vol. 24, no. 4, 1992, pp. 325-376.

- [35] Maintz B. A. and Viergever M. A., “A Survey of Medical Image Registration”, *Medical Image Analysis*, vol. 2, no. 1, Apr. 1998, pp. 1-36.
- [36] T. M. Lehmann, H. G. Gröndahl, and D. K. Benn, “Computer-based registration for digital subtraction in dental radiology”, *Dentomaxillofacial Radiology*, vol.29, issue. 6, Nov. 2000, pp. 323-346.
- [37] J. L. Ostuni, E. Fisher, S. M. Dunn, and P. van de Stelt, “Geometric Nature of Dental Radiograph Formation”. *Proc. Of the 15th Annual Int’l conference of the IEEE on Engineering in Medicine and Biology Society* 1993
- [38] D. C. Yoon, “A new method for the automated alignment of dental radiographs for digital subtraction radiography”, *Dentomaxillofacial Radiology*, vol 29, issue 1, Jan. 2000 pp. 11-19.
- [39] E. I. Zacharaki, G. K. Matsopoulos, P. A. Asvestas, K. S. Nikita, K. Gröndahl and H. G. Gröndahl, “A digital subtraction radiography scheme based on automatic multiresolution registration”, *Dentomaxillofacial Radiology*, vol. 33, issue 6, Nov. 2004, pp. 379-390.
- [40] F. Samadzadegan, F. Bashizadeh, M. Hahn, P. Ramzi, “Automatic Registration of Dental Radiograms”, *Proc. of Geo-Imagery Bridging Continents, Commission 5, XXth ISPRS Congress*, Jul. 2004, pp. 328 - 331, Istanbul – Turkey.
- [41] A. K. Jain, R. Duin, and J. Mao, “Statistical Pattern Recognition: A Review”, *IEEE Trans. PAMI*, vol. 22, no. 1, 2000, pp. 4-37.
- [42] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. John Wiley & Sons, Inc. Second Edition 2001
- [43] G.V. Trunk, “A Problem of Dimensionality: A Simple Example”. *IEEE Transaction on PAMI*, vol. 1, no. 3, Jul. 1979, pp. 306-307.
- [44] F.J. Ferri, P. Pudil, M. Hatel, and J. Kittler, “Comparative Study of Techniques for Large-Scale Feature Selection”. In *Pattern Recognition in Practice IV* (Gelsema and Kanal, eds.), Elsevier, 1994, 403-413.
- [45] Matthew Turk and Alex Pentland, “Eigenfaces for Recognition”, *Journal of Cognitive Neurosciences*, vol. 3, no. 1, 1991, pp. 71-86.
- [46] Matthew A. Turk and Alex P. Pentland, “Face Recognition Using Eigenfaces”, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586-591.

- [47] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection". *IEEE Transactions on PAMI*, vol. 19, no. 7, Jul. 1997, pp. 711-720.
- [48] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang, "Face Recognition Using Laplacianfaces", *IEEE Transactions on PAMI*, vol. 23, no. 3, Mar. 2005, pp. 328-340.
- [49] Josef Kittler, Mohamad Hatef, Robert Duin, and Jiri Matas, "On Combining Classifiers", *IEEE Trans. on PAMI*, vol. 20, no. 3, Mar. 1998, pp. 226-239.
- [50] David M.J. Tax, Martijn van Breukelen, Robert P. W. Duin, and Josef Kittler, "Combining multiple classifiers by averaging or by multiplying?", *Pattern Recognition*, vol. 33, no. 7, Jul. 2000, pp. 1475-1485.
- [51] Amanda Sharkey, and Noel Sharkey, "How to Improve the Reliability of Artificial Neural Networks", *Technical Report CS-95-11*, Department of Computer Science, University of Sheffield 1995.
- [52] Anil Jain, Karthik Nandakumar, and Arun Ross "Score Normalization in Multimodal Biometric Systems", *Pattern Recognition*, 38(2005), pp. 2270-2285.
- [53] C. Sanderson, and K. K. Paliwal, "On the Use of Speech and Face Information for Identity Verification", *IDIAP Research Report*, 04-10, Martigny, Switzerland 2004.
- [54] Arun Ross, and Anil Jain, "Information fusion in biometrics", *Pattern Recognition Letters*, vol. 24, issue 13, Sep. 2003, pp. 2115-2125.
- [55] J. Kittler, Y. P. Li, J. Matas, and M. U. Sanchez, "Combining Evidence in Multimodal Personal Identity Recognition Systems", in *Proc. 1st Int. Conf. on AVBPA*, Crans-Montana, Switzerland, Mar. 1997, pp. 327-334
- [56] A. K. Jain, S. Prabhakar, and A. Ross, "Fingerprint matching: Data acquisition and performance evaluation", *Technical Report MSU-TR: 99-14*, Michigan State University 1999.
- [57] Anil Jain, and Sharath Pankanti, "Biometric Systems: Anatomy of Performance", *IEICE Trans. Fundamentals*, special issue on Biometrics vol. E84-D, no. 7, pp. 788-799, 2001.
- [58] John Daugman, "Biometric decision landscapes", *Technical Report: TR482*, University of Cambridge Computer Lab, <http://www.cl.cam.ac.uk/users/jdg1000>, 1999.

- [59] John Daugman, "Recognizing persons by their iris patterns", in *Biometrics: Personal Identification in a Networked Society*, A.K. Jain, R. Bolle and S. Pankanti (eds), Kluwer Academic Publishers, NY 1999.
- [60] S. Pankanti, S. Prabhakar, and A. K. Jain, "On the Individuality of Fingerprints", *IEEE Trans. on PAMI*, vol. 24, no. 8, Aug. 2002, pp. 1010-1025.
- [61] D. A. Stoney and J. I. Thornton, "A critical analysis of quantitative fingerprint individuality models", *Journal of Forensic Sciences*, vol. 31, no. 4, 1187-1216, 1986.
- [62] N. Ratha, J. Connell, R. Bolle. "Enhancing Security And Privacy In Biometrics-Based Authentication Systems", *IBM Systems Journal*, vol. 40, no. 3, 2001, pp. 614-634.
- [63] J. Canny, "A Computational Approach to Edge Detection", *IEEE Transactions on PAMI*, vol. 8, no. 6, 1986, pp. 679-698.
- [64] CJIS Division – Washington State Patrol Missing and Unidentified Person Unit, Digitized Dental Images (Database), May 2000
- [65] CJIS Division – ADIS project, Digitized Radiographic Images (Database), Aug. 2002
- [66] Dr. Robert Howell, Digitized Radiographic Images (Synthetic Database), Nov. 2004
- [67] Faisal Chaudhry, "Performance Evaluation of the Decision-Making stage in Automated Dental Identification System using Receiver Operating Characteristics curve and Bootstrap methods", *Masters Problem Report*, Department of industrial and Management Systems Engineering, West Virginia University, Apr. 2005.
- [68] J. Daugman, "High Confidence Visual Recognition of Persons by a Test of Statistical Independence", *IEEE Transactions on PAMI*, vol. 15, no. 11, Nov. 1993, pp. 1148-1161.
- [69] Diao Eldin M. Nassar, and Hany H. Ammar, "A Neural Network System for Matching Dental Radiographs", submitted to *Pattern Recognition*.
- [70] Roger B. Nelsen, *An Introduction to Copulas*, Lecture Notes in Statistics, vol. 139. Springer-Verlag Telos 1999.
- [71] Dominique Drouet Mari, and Samuel Kotz, *Correlation and Dependence*. Imperial College Press 2001.
- [72] Norman Abramson, *Information Theory and Coding*. McGraw-Hill, New York 1963.
- [73] Thomas M. Cover, and Joy A. Thomas, *Elements of Information Theory*. John Wiley, 1991.

- [74] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, Addison Wesley, 1993
- [75] P. Baldi, and Y. Chauvin, "Neural Networks for Fingerprint Image Recognition", *Neural Computations*, vol. 5, 1993
- [76] H. H. Ammar, S. Zeng, and Z. Miao, "Parallel Processing and Fingerprint Image Comparison". *International Journal of Modeling and Simulation*, vol. 18, no. 2, pp. 85-99, 1998
- [77] H. H. Ammar, and Z. Miao, "Parallel Algorithms for a Neural Network based Automated Fingerprint Image Comparison System". *The International Journal of Supercomputer Applications and High Performance Computing*, vol. 14, no. 1, pp. pp. 3-25, January 2000.
- [78] S. M. Ali and S. D. Silvey, "A General Class of Coefficients of Divergence of One Distribution from Another", *J. Royal Statistics Soc.*, vol. 28, pp. 132-142, 1966.
- [79] J. Shore, and R. Johnson, "Properties of cross-entropy minimization", *IEEE Trans. on Information Theory*, vol. 27, no. 4, pp. 472-482, July 1981.
- [80] C. Looney, *Pattern Recognition Using Neural Networks*. Oxford University Press 1997.
- [81] L. Fausett, *Fundamentals of Neural Networks – Architectures, Algorithms, and Applications*. Prentice Hall 1994.
- [82] S. White, and M. Pharoah, *Oral Radiology Principles and Interpretation*. Mosby, Inc. Fourth Edition 2000.