Reconstruction of source images from quantized biometric match score data

Andy Adler
School of Information Technology and Engineering, University of Ottawa, Ontario, Canada
adler@site.uottawa.ca

1. Introduction

One important consideration for the privacy and security of biometric systems is whether the enrolled images may be regenerated from template or match score information. Many vendors of biometric systems have claimed that it is impossible or infeasible to recreate the image from the templates; in light of this claim, biometric data has been considered non-identifiable, and managed in ways that the source images cannot. For example, this assumed non-identifiability has been used to allay concerns that fingerprint, face, and iris images may accessed from their storage on identification cards.

Several authors have shown that it is possible to reconstruct a good estimate of an unknown enrolled image from a fingerprint or face recognition template [1,5,8]. In each case, an algorithm is developed which allows small, but physiologically realistic, modifications to be made to an input image. These modifications are used to perform a "hill-climbing" attack. The modified image is presented to the algorithm and compared against the enrolled image to obtain a match score. Modifications that increase the match score are retained. Eventually, a best-match image is generated; this image resembles the essential features of the unknown enrolled image, and is able to compare to it at high match score. The implication is that an attacker has effective access to the enrolled biometric image, which, depending on the details of the security system, may have implications in terms of the system privacy and security.

In order to protect against this attack, the BioAPI [4] specifies that match scores should be quantized. A biometric system which implements this API would round the calculated match score to the nearest value in a preselected sequence. This quantization would normally prevent a hill-climbing algorithm from seeing any modification in match score due to small modifications in the input image. However, in this paper, a modification of the hill-climbing algorithm is described to allow it to work successfully in with such quantized match scores [2].

2. Modified hill-climbing algorithm

This section describes the modified hill-climbing algorithm to regenerate an enrolled image (IM_{targ}). The algorithm iteratively updates its estimate (IM_i) using quantized match score data represented by the biometric comparison function *match_score*( ).

- *Local database preparation:* A local database (*LD*) of frontal pose face images is obtained. Images are rotated, scaled, and cropped such that all images have the same size and eye locations.
- *Eigenface calculation:* An eigenimage (*EF*) decomposition of *LD* is calculated. The image is then divided into quadrants. Quadrant eigenimages ($EF_{k,quadrant}$) are defined to be equal to EF_k within the quadrant and zero elsewhere.
- *Initial image selection:* A selection of images is chosen randomly from *LD* and individually compared to the target. The initial estimate (IM_0) is selected to be the one with the highest match score.
- *Iterative estimate improvement:* Iterate the following steps (for step number *i*).
  - Randomly select an eigenimage, EF_k
  - Randomly select a quadrant *Q*. The diametrically opposite quadrant is *OQ*.
  - Generate an image *RN* consisting of random Gaussian noise in *OQ* and zero elsewhere.
  - Calculate the required contribution of *RN* to reduce the match score by one quantization level.
  - Iterate for step number *j* for a small range of values c_j. Calculate: $MS_j = match\_score(NI + c_j \times EF_{k,Q}, IM_{targ})$
  - Select j_{max} as the value of *j* for which MS_j is maximized.
  - Calculate $IM_{i+1} = IM_i + c_{jmax} \times EF_{k,Q}$
  - Truncate pixel values to image limits (ie. 0 to 255) if any pixel values of IM_{i+1} exceed the limits.
- Repeat iterations until there is no significant improvement in match score.

Values of c_j were selected heuristically for fastest convergence; the maximum value of *c* represented approximately 10% of the standard deviation of target image pixel values. 3000 iterations and 15 values of c_j (including zero) were used. This algorithm works separately on quadrants of the image; because the quantized match score will not normally give information to allow hill-climbing, a carefully chosen level of noise is introduced into the opposite image quadrant, in order to force the quantized score into a range where its information can once again be used.

The modified hill-climbing algorithm was applied to a commercially available face recognition software package, modified to provide quantized match score output. Three quantization levels, 1.0, 0.5 and 0.001, were tested. The

latter level is effectively the same as no quantization. The algorithm required 135,000 biometric comparisons, and took 122 minutes on a 2.8 GHz Pentium IV PC computer. The 100 lowest order eigenimages were used for image regeneration. Match score results are shown in terms of the *confidence* of a genuine match, where the *confidence* is the likelihood that a comparison was genuine, given a match score *MS* was obtained [3]. Figure 1 shows a graph of *confidence* versus iteration number for a representative image at each quantization level.

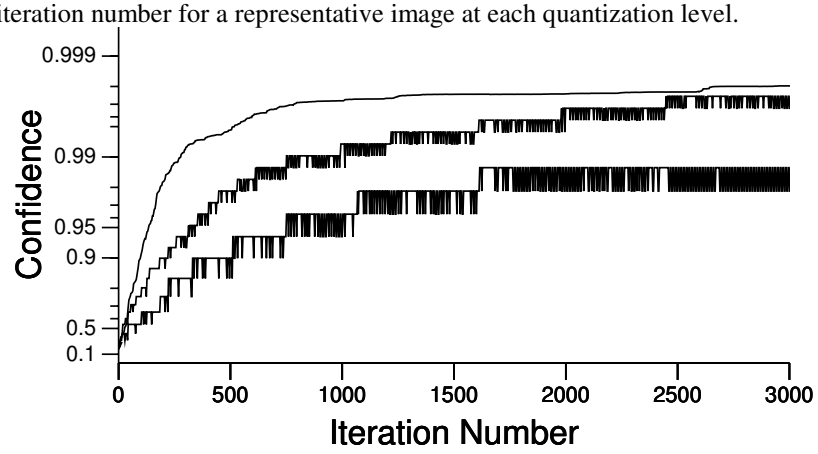

Figure 1: *Confidence* (calculated from match score value) versus iteration number for the modified hill-climbing algorithm for a representative initial image (Top curve, quantization = 0.001; Middle Curve, quantization = 0.5; Bottom Curve, quantization = 1.0). Increasing quantization level decreases the maximum *confidence* achieved by the algorithm.

3. Discussion

This paper shows that it is possible to regenerate biometric source images from match score data even if the BioAPI [4] recommendation on match score quantization is implemented. The hill-climbing algorithm is augmented in order to modify the test image at each iteration such that the quantized match score provides useful information. This algorithm was tested for three different quantization levels, and shown to be able to successfully regenerate an image which verifies at a high *confidence* to the original. As the quantization level increased, the algorithm running time increased and the maximum *confidence* obtained decreased. With a quantization level corresponding to a 10.6% change in *confidence*, the calculated image would not be able to successfully masquerade against a system with a false accept rate setting of 1.0%. On the other hand, such a severe setting for the quantization level removes a significant amount of information from the calculated match score values. Furthermore, we anticipate that significant improvements are possible to this algorithm. We conclude that the quantization of match score values does not protect against the regeneration of images from stored biometric templates. This result has privacy and security implications for the storage and transmission of biometric data, including biometric templates and match score values.

This technique may be extended for systems using biometric encryption (e.g. [6,7]). In such systems, the template does not contain a representation of the biometric image, but rather a security token "encrypted" with the image. In order for such algorithms to allow for variability in the input image, the token must be robustly encoded, using an error correcting code. In general, a match score can be constructed from an error correcting code as the distance to the nearest valid code word. Given a definition of a match score, it would be possible to use hill-climbing techniques to regenerate the target image, and then use this image to obtain the security token.

4. References

1. Adler A, (2003) Sample images can be independently restored from face recognition templates, *Proc. Can. Conf. Elec. Comp. Eng.* Montreal, Canada, May 2003. 1163-1166
2. Adler A, (2004) Images can be regenerated from quantized biometric match score data, *Proc. Can. Conf. Elec. Comp. Eng.*, Niagara Falls, Canada, May 2004.
3. Bazen A M, Veldhuis R N J (2004) Likelihood-ratio-based biometric verification, *IEEE Trans. Circuits Systems Video Technol.* **14** 86-94
4. BioAPI Consortium (2001) *BioAPI Specification (Version 1.1)* http://www.bioapi.org/BIOAPI1.1.pdf (current May 2004)
5. Hill C J (2001), *Risk of Masquerade Arising from the Storage of Biometrics*, B.S. Thesis, Australian National University http://chris.fornax.net/biometrics.html (current May 2004)
6. Monrose F, Reiter M K, Li Q, Wetzel S, (2001) Cryptographic key generation from voice, *Proc. IEEE Conf. Security and Privacy.*
7. Soutar C, Roberge D, Stoianov A, Gilroy R, Vijaya Kumar B V K, (1998) Biometric Encryption using image processing, *Proc. SPIE* 3314:178-188.
8. Soutar C, Gilroy R, Stoianov A, (1999) Biometric System Performance and Security *Conf. IEEE Auto. Identification Advanced Technol.*