ROBUST FEATURE-LEVEL MULTIBIOMETRIC CLASSIFICATION

Ajita Rattani[1], D. R. Kisku[1], Manuele Bicego[2] and Massimo Tistarelli[2]
[1]Department of Computer Science and Enginnering,
Indian Institute of Technology Kanpur,
India
[2]Computer Vision Laboratory,
University of Sassari,
Italy

ABSTRACT

This paper proposes a robust feature level based fusion classifier for face and fingerprint biometrics. The proposed system fuses the two traits at feature extraction level by first making the feature sets compatible for concatenation and then reducing the feature sets to handle the 'problem of Curse of Dimensionality'; finally the concatenated feature vectors are matched. The system is tested on the database of 50 chimeric users with five samples per trait per person. The results are compared with the monomodal ones and with the fusion at matching score level using the most popular sum rule technique. The system reports an accuracy of 97.41% with a FAR and FRR of 1.98% and 3.18% respectively, outperforming single modalities and score-level fusion.

1. INTRODUCTION

In recent years, biometrics authentication has seen considerable improvements in reliability and accuracy, with some of the traits offering good performance. However, even the best biometric traits to date are facing numerous problems, some of them inherent to the technology itself. Thus a single biometric is not sufficient to meet the variety of requirements including matching performance imposed by several large-scale authentication systems.

Multibiometric systems [1] remove some of the drawbacks of the uni-biometric systems by grouping the multiple sources of information. They address the problem of non-universality, since multiple traits provide sufficient population coverage. They also limit spoofing since it would be difficult for an impostor to spoof multiple biometric traits/ information of a genuine user simultaneously [2]. Ross and Jain [3] have presented an overview of Multimodal
Biometrics and have proposed various levels of fusion, various possible scenarios, different modes of operation, integration strategies and design issues.

Evidence in a multi-biometrics system can be integrated in several different levels as described below:

1. **Sensor level**: The raw data acquired from multiple sensors can be processed and integrated to generate new data from which features can be extracted. For example, in the case of fingerprint biometrics, the fingerprint image acquired from both optical and solid state sensors may be fused to generate a single image which could then be subjected to feature extraction and matching.
2. **Feature level**: Information extracted from the different sources is concatenated into a joint feature vector, which is then compared to an enrollment template (which itself is a joint feature vector stored in a database) and assigned a matching score as in a single biometric system
3. **Match score level:** Feature vectors are created independently for each modality and are then compared to the enrollment templates which are stored separately for each biometric trait. Based on the proximity of feature vector and template, each subsystem computes its own matching score. These individual scores are finally combined into a total score, which is passed to the decision module.
4. **Rank level:** This type of fusion is relevant in identification systems where each classifier associates a rank with every enrolled identity. Thus, fusion entails consolidating the multiple ranks associated with an identity and determining a new rank that would aid in establishing the final decision.
5. **Decision level:** A separate authentication decision is made for each biometric trait. These decisions are then combined into a final vote. Fusion at the decision level is considered to be rigid due to the availability of limited information

The Biometric system that integrates information at an earlier stage of processing is expected to provide more promising results than the systems that integrate information at later stage because of availability of more/ richer information. Since the feature set contains richer information about the input biometric data than the matching score or the output decision of a matcher, fusion at the feature level is expected to provide better recognition performance.

1-4244-0487-8/06/$20.00 ©2006 IEEE

Fusion at the match score, rank and decision levels have been extensively studied in the literature. As early as 1993, Chibelushi et al. have proposed in [4] to integrate acoustic and visual speech (motion of visible articulators) for speaker recognition, using a simple linear combination scheme. Duc et al. proposed in [5] a simple averaging technique and compared it with the Bayesian integration scheme presented by Big¨un et al. Kittler et al. have proposed in [6] a multimodal person verification system, using three experts: frontal face, face profile, and voice. The best combination results are obtained for a simple sum rule. Hong and Jain have proposed in [7] a multi-modal personal identification system which integrates two different biometrics (face and fingerprints) that complement each other.

However, fusion at the feature level is a relatively understudied problem. Only work of Ross and Govindarajan [8] is reported in the literature for the fusion of hand and face biometrics at feature extraction level. Fusion at this level is difficult to achieve in practice because multiple modalities may have incompatible feature set or the feature space may be unknown, concatenated feature vector may lead to the problem of curse of dimensionality, a more complex matcher may be required for concatenated feature vector and concatenated feature vector may contain noisy or redundant data thus leading to decrease in the performance of the classifier [8].

This paper proposes a robust feature level based fusion classifier which integrates face based on SIFT features and fingerprint based on minutiae matching at feature extraction level. First the feature set extracted from two traits are made compatible for concatenation then feature reduction is done to handle the 'problem of curse of dimensionality' [9]; finally the matching of the concatenated feature vector is determined. The results are compared with the fusion at matching score level using the most popular sum rule technique. This work reports high increase in the performance of the system as compared to fusion at matching score level.

The rest of this paper is as follows: section 2 briefly describes the face and fingerprint algorithms together with the modifications made to enable the concatenation at feature extraction level. Section 3 describes the proposed fusion strategy. Experimental results are given in section 4 and in the last section the conclusions are drawn.

2. FACE AND FINGERPRINT ALGORITHMS

2.1. Face Recognition based on SIFT Features

Face Recognition is a noninvasive process where a portion of the subject's face is photographed and the resulting image is reduced to a digital code. A face recognition system is developed based on SIFT features [10]. These features are invariant to image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projection. Features are efficiently detected through a staged filtering approach that identifies stable points in scale space and are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images.

Following are the major stages of computation used to generate the set of image features:

1. **Scale-space extrema detection:** The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.
2. **Keypoint localization:** At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.
3. **Orientation assignment:** One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.
4. **Keypoint descriptor:** The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant required levels of local shape distortion and change in illumination [10].

Due to the stability and robustness of these features, they have been recently applied to face recognition problem [11] (Figure 1). Thus the input to the system is the face image and the output is the set of extracted SIFT features $s=(s_1, s_2, \ldots\ldots s_m)$ where each feature $s_i=(x, y, \theta, Keydesc)$ consist of x, y spatial location, θ as local orientation and keydescriptor of size 1x128.

Previous work [11] only considered the local keypoint descriptor extracted at SIFT locations for verifying the proximity between the database and query image. The current implementation of the system employs spatial coordinates and local orientation along with the keydescriptor for the authentication purposes. The system has been tested on a part of the BANCA database [11, 12] (see Section 4 for details) giving us the accuracy of 88.9% (for the definition of accuracy refer to section 4).

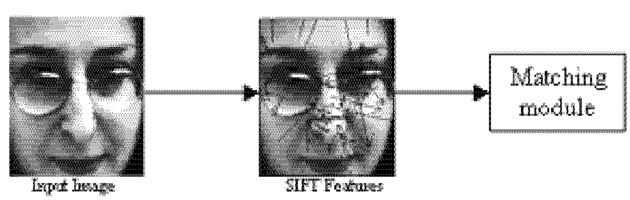

Figure 1: Example image used for face recognition using the SIFT Features

2.2. Fingerprint Verification based on minutiae

1-4244-0487-8/06/$20.00 ©2006 IEEE

Fingerprints are the most widely used biometric feature for person identification and verification. Fingerprints encompass two main types of features that are used for automatic fingerprint identification and verification: (i) global ridge and furrow structure that forms a special pattern in the central region of the fingerprint and (ii) minutiae details associated with the local ridge and furrow structure. The fingerprint recognition module has been developed using minutiae based technique [13] as shown in Figure 2.

2.2.1. Image Segmentation and Rotation Invariance

The input image is segmented to remove noise and extract the inked region (the foreground part) from the background region. The image is also transformed in order to obtain rotation invariance, by detecting the left, top and right edges of the foreground to calculate the overall slope of the foreground, and by fitting a straight line to each edge by linear regression. The left and right edges, which are expected to be roughly vertical, use lines of the form $x = my + b$ and the top edge use the form $y = mx + b$. The overall slope is defined to be the average of the slopes of the left-edge line, the right-edge line, and a line perpendicular to the top edge line. A rectangle is fitted to the segmented region and rotated with the same angle to nullify the effect of rotation.

2.2.2. Image Enhancement

The segmented image is enhanced using local ridge orientation at [x, y] which is the angle θ_{xy} that the fingerprint ridges, crossing through an arbitrary small neighborhood centered at [x,y], form with the horizontal axis. The local frequency f_{xy} at point [x, y] is the inverse of the number of ridges per unit length along a hypothetical segment centered at [x,y] and orthogonal to the local ridge orientation θ_{xy}. Gabor filter is tuned to the local ridge orientation and the local ridge frequency to get the enhanced image. This is followed by binarization and thinning.

2.2.3. Minutiae Extraction

Minutiae extraction was carried out using the crossing number approach. Crossing number of pixel '*p*' is defined as half the sum of the differences between pairs of adjacent pixels defining the 8-neighborhood of '*p*'. A minutiae *m* is described by the triplet $m=\{x, y, \theta\}$, where *x, y* indicate the minutiae location coordinates and θ denotes the minutiae orientation, which is the orientation evaluated for the minutiae location from the orientation image obtained during the enhancement process. Thus the input to the system is the fingerprint image and output is set of minutia $m=(m_1, m_2, \ldots \ldots m_m)$.

The SIFT feature set is a translation and rotation invariant set, composed by of the keydescriptor along with the spatial location and orientation. In the proposed approach, minutiae feature set is made compatible with the SIFT feature set. The smooth flow pattern of ridges and valleys in a fingerprint can be viewed as an oriented texture field [14]. Textured regions possessing different spatial frequency, orientation, or phase can be easily discriminated by decomposing the texture in several spatial frequency and orientation channels. The local region around each minutiae point is convolved with the bank of gabor filters to analyze local texture information for eight different degrees of orientation *(0 , 22.5 , 45 , 67.5 , 90 , 112.5 ,135 , and 157.5)*, eight different scales and two phases thus giving 1x128 keydescriptor. The rotation invariance is handled during the preprocessing step and the translation invariance is handled by registering the database image with the query images using reference point location [14]. Scale invariance is not a significant problem since most fingerprint images can be scaled as per the dpi specification of the sensors. This makes the feature set of SIFT based face recognition compatible with the minutiae matching as explained in experimental section. This modified algorithm has been tested on a proprietary Database (see sect. 4 for details), giving an accuracy of 91.82%.

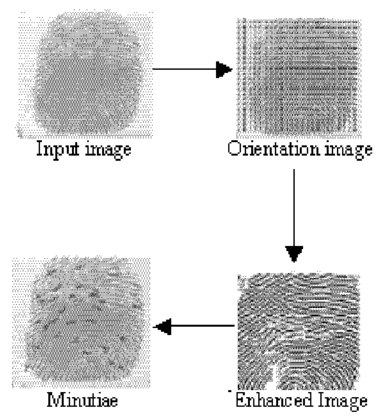

Figure 2: Preprocessing steps of the fingerprint verification based on minutiae

3. FEATURE LEVEL FUSION STRATEGY

The feature level fusion is a simple concatenation of the feature sets obtained from different sources of information. Let $X = (x_1, x_2, \ldots x_m)$ and $Y = (y_1, y_2, \ldots \ldots y_m)$ denote feature vectors representing the information extracted by two different sources. Vector Z is formed by concatenation of these two feature sets, which would have better recognition capability of the individual. Thus vector *Z* is generated by first augmenting vectors *X* and *Y*, normalizing the feature vectors to ensure the same range and scales of values and then performing feature selection/reduction techniques on the resultant feature vector set. The vector *Z* is then input to the matcher which computes the proximity between two concatenated feature vectors [8].

3.1. Features normalization and Concatenation

1-4244-0487-8/06/$20.00 ©2006 IEEE

The extracted features from face recognition using SIFT $s=(s_1, s_2, \ldots\ldots s_m)$ and fingerprint based minutiae $m=(m_1, m_2, \ldots\ldots m_m)$ are first normalized to ensure the same scale and range for both the feature vectors and to enable the compatibility between the two feature sets. The "*min-max*" normalization technique [15] is used to normalize the keydescriptors. This normalization technique results in the mapping of features values to the range [0 to 1].

Let s_{norm} and m_{norm} represents the normalized feature sets of face and fingerprint. These features are then concatenated into a single feature set as $concat=(s_{1norm}, s_{2norm}, \ldots s_{mnorm}, \ldots m_{1norm}, m_{2norm}, m_{mnorm})$.

3.2. Feature Reduction and Matching

Concatenated feature vector *concat* belong to R^{m+n}. The `curse-of-dimensionality' related to feature level fusion states that the concatenated feature vector need not necessarily improve the matching performance of the system as some of the feature values may be noisy and redundant compared to the others. Thus the feature selection/ reduction is applied to get the optimal subset of features of size k, $k<(m+n)$ that improves the performance of the classifier. The redundant features in the proposed system are removed using "*K-means*" clustering techniques [16] and choosing the most proximate feature to the mean of the cluster as the representative of the set of similar features. The optimal features are matched using the point pattern matching algorithm where the pair of points is considered matching only if the spatial distance, direction distance and the euclidean distance between the corresponding key descriptors are within some threshold where each point in query $concat`_j$ and database feature set $concat_i$ contain (x, y, ϴ, keydesc). Thus a point $concat`_j$ in input set is considered matching with the template set $concat_i$ if the spatial distance (sd) between them is smaller than a given tolerance r_0, the direction difference (dd) between them is smaller than an angular tolerance Θ_o and euclidean distance(euc) (equation 1 and 2) between the keydescriptor is between some threshold:

$$sd(concat'_j, concat_i) = \sqrt{(x'_j - x_i)^2 + (y'_j - y_t)^2} \le r_0 \quad \textbf{1}$$

$$dd(concat'_j, concat_t) = \min(|\theta'_j - \theta_t|, 360^o - |\theta'_j - \theta_t|) \le \theta_o \quad \textbf{2}$$

The final matching score is computed on the basis of number of matched pairs found in the two sets. Figure 3 represents the fusion classifier.

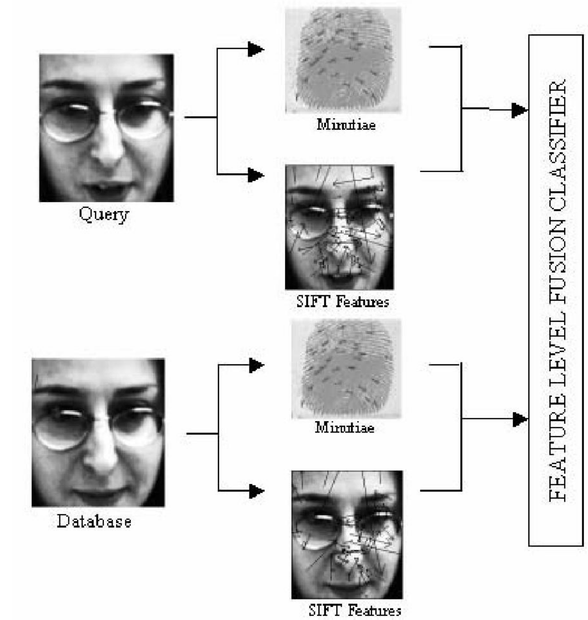

Figure 3: Scheme of the feature level based Fusion Classifier

4. EXPERIMENTAL RESULTS

The Database used for testing consists of 50 chimeric individuals composed of 5 face and fingerprint images for each individual keeping in mind the independence of face and fingerprint traits. The face images are taken from the controlled session of BANCA Database and fingerprint images are collected by the authors for this experimental purpose. The fingerprint images are acquired using an optical sensor at 500 dpi.

The following training and testing procedure has been established for mono-modals and multimodal system:

Training: one image per person is used for enrollment in the face and fingerprint verification system; for each individual, one pair face-fingerprint is used for training the fusion classifier.

Testing: four samples per person are used for testing and generating client scores. Impostor scores are generated by testing the client against the first sample of the rest of the individuals, in the case of monomodal systems. In case of multimodal testing the client is tested against the first face and fingerprint samples of the rest of the chimeric users thus in total 50x4=200 client scores and 50x49=2450 imposters scores for each of uni-modal and multimodal system are generated.

Experiments were conducted in two sessions. In the first experiment, the unimodal systems were modified to enable the feature level fusion. The matching module of the face recognition based on SIFT Features was modified to include the spatial and orientation features along with the keydescriptor as the part of the extracted features. The system has been tested using part of the BANCA Database.

Both the FAR and FRR were computed varying the acceptance threshold. The accuracy is established setting verification threshold corresponding to minimal value of both FAR and FRR. The computed accuracy of the system is 88.9% accurate with a FAR and FRR of 10.52% and 11.47% respectively at a threshold of 65. The feature extraction module for fingerprint recognition module was modified to include the local keydescriptor along with the spatial and orientation information. The system has been tested using the above mentioned protocol and found to be 91.82% accurate with a FAR and FRR of 10.97% and 5.38% respectively at a threshold of 45.

In the next experiment, face and fingerprint classifiers were combined at matching score level using the most popular sum of scores technique. The system is found to be 94.77% accurate with a FAR and FRR of 4.78% and 5.66% respectively at a threshold of 50. Finally these two traits are combined at feature extraction level and tested using the above mentioned protocol with the accuracy of 97.41% and FAR and FRR of 1.98% and 3.18% respectively. Thresholds of different systems are fixed by analyzing the results obtained at different thresholds. Infact FAR-FRR were only used to determine the optimal thresholds. Table 1 shows the FRR, FAR and accuracy of the monomodal and multimodal system at matching score and feature extraction level fusion. Figure 4 shows the accuracy graph of various systems where the x axis represents the threshold values and y axis represents the accuracies and the ROC curves are shown in Figure 5 where x axis represents the FRR (%) and y axis represents FAR(%).

The obtained results demonstrate the performance superiority of the feature level based fusion classifier in comparison with the matching score level based classifier and the monomodal systems.

Table 1: FRR , FAR and Accuracy values

Algorithm	**FRR(%)**	**FAR(%)**	**Accuracy**
Face SIFT	11.47	10.52	88.90
Fingerprint	5.384	10.97	91.82
Face+Finger at Matching score level	5.66	4.78	94.77
Face+Finger at Feature Extraction Level	1.98	3.18	97.41

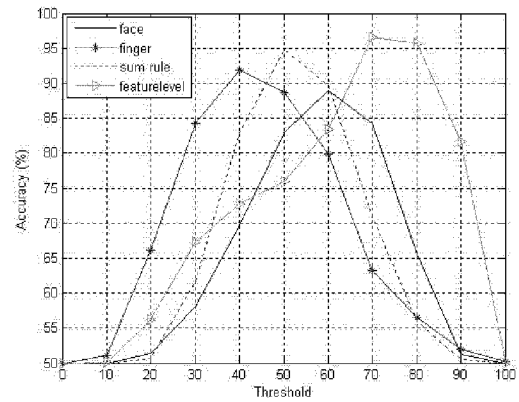

Figure 4: Computed accuracy as function of the verification threshold. Individual face and fingerprint modalities and both the score-level and feature-level results are shown

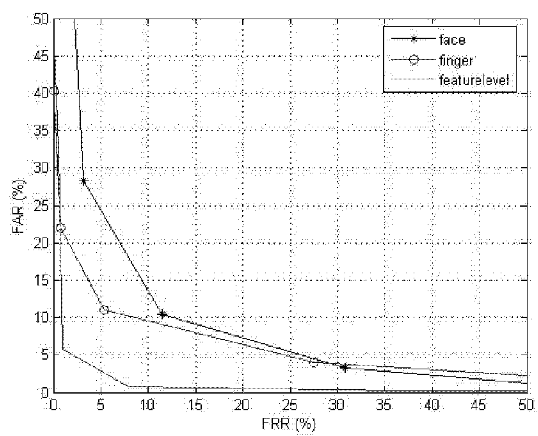

Figure 5: Computed ROC curves (FAR vs FRR) for the single modalities and the feature level fusion.

5. CONCLUSION

A multimodal biometric system based on the integration of face and a fingerprint trait was presented. These two traits are the most widely accepted biometrics; moreover there are other advantages in such a system including the easy of use nd the availability of low-cost, off-the-shelf hardware for data acquisition.

From a system point of view, redundancy can be always exploited to improve accuracy and robustness. This is achieved in many living systems as well. Human beings, for example, use several perception cues for the recognition of other living creatures. They include visual, acoustic and tactile perception. Starting from these considerations, this paper also outlined the possibility to augment the verification accuracy by integrating the fingerprint and face biometric traits. In most of the examples presented in the literature, fusion is performed either at the score level or at the decision level, always improving the performance of each single modality.

In this paper a novel approach has been presented were both fingerprint and face images are processed with compatible feature extraction algorithms to obtain comparable features from the raw data.

The performances of both the monomodal verification and the score level fusion were compared against the feature level fusion. Interestingly the feature level fusion provided far better results than any of the other cases, also outperforming of more than 2% the score level fusion.

These results, not only confirm the validity of the multimodal or multibiometrics approach, but also enforce the need for the definition of compatible processing channels for each biometric trait.

6. REFERENCES

[1] L. Hong, A. Jain and S. Pankanti, "Can Multibiometrics Improve performance", Proc. of AutoID 99, pp. 59-64, 1999.

[2] A. K. Jain and A. Ross, "Multibiometric systems", Communications of the ACM, 47 (1), pp. 34-40, 2004.

[3] A. Ross and A.K. Jain, "Information Fusion in Biometrics", Pattern Recognition Letters, 24, pp. 2115-2125, 2003.

[4] C.C. Chibelushi, J.S. Mason, and F. Deravi, "Integration of acoustic and visual speech for speaker recognition", In EUROSPEECH'93, pp. 157–160, 1993.

[5] B. Duc, G. Maˆıtre, S. Fischer, and J. Big¨un, "Person authentication by fusing face and speech information", In Proc. of the First International Conference on Audio and Video-based Biometric Person Authentication, Lecture Notes in Computer Science. Springer Verlag, 1997.

[6] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers", IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3), pp. 226–239, March 1998.

[7] L. Hong and A. Jain, "Integrating Faces and Fingerprints for Personal Identification", IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(12), pp. 1295–1307, December 1998.

[8] A. Ross and R. Govindarajan, "Feature Level Fusion Using Hand and Face Biometrics", Proc. of SPIE Conference on Biometric Technology for Human Identification II, Orlando, USA, pp. 196-204, March 2005.

[9] G. Trunk, "A problem of dimensionality: A simple example", IEEE Trans. Pattern Analysis and Machine Intelligence, 1(3), pp. 306-307, 1979.

[10] Lowe and G. David, "Object recognition from local scale invariant features," International Conference on Computer Vision, Corfu, Greece, pp. 1150–1157, September 1999.

[11] M. Bicego, A. Lagorio, E. Grosso and M. Tistarelli, "On the use of SIFT features for face authentication", Proc. of Int Workshop on Biometrics, in association with CVPR 2006.

[12] E. Bailly-Baillire, S.Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Marithoz, J. Matas, K. Messer, V. Popovici, F. Pore, B. Ruiz, and J. P Thiran, "The BANCA database and evaluation protocol", In Proc. Int. Conf on Audio- and Video-Based Biometric Person Authentication (AVBPA03), Springer-Verlag , pp. 625-638, 2003.

[13] N. K. Ratha, K. Karu, S. Chen, and A. K. Jain, "A Real-time Matching System for Large Fingerprint Databases", IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8), pp. 799-813, 1996.

[14] A. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "Filterbank-based fingerprint matching," IEEE Transactions on Image Processing, vol. 9, no. 5, pp. 846- 859, 2000.

[15] A. Jain, K. Nandakumar and A. Ross, "Score Normalization in Multimodal Biometric Systems", to appear in Pattern Recognition, 2005.

[16] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. Piatko, R. Silverman and A. Y. Wu, "The Analysis of a simple k-means Algorithm", Technical Report CAR-TR-937, Center for Automation Research, University of Maryland, College Park, January 2000.

1-4244-0487-8/06/$20.00 ©2006 IEEE