Comparison Study of Gaussian Mixture Models for Fingerprint Image Duplication with a New Model

Sos S. Agaian,

Division of Engineering, The University of Texas at San Antonio. One UTSA Circle, San Antonio, TX-782549

Rushikesh D. Yeole, The University of Texas at San Antonio. One UTSA Circle, San Antonio, TX-78249

Gary Reinecke Boston University

Mary-Ann FlashScan3D

Mike Troy FlashScan3D

Abstract: This paper presents a comparison study of Gaussian Mixture Models for fingerprints image duplication and analysis. It also presents a new probabilistic Parametric Gaussian Mixture Model(GMM). The system is built around the likelihood ratio test for verification, using simple but effective GMMs for likelihood functions and a form of Bayesian adaptation to derive the models. The Computer simulation show that the developed new algorithms have the most optimal performance as compared to state of art algorithms GMMs, Generalized GMMs, Finite Bayesian learning for GMMS, Texture Synthesis and Improved Adaptive Algorithm. The performance of the presented algorithm was evaluated by Bovik Index, Entropy and Mean Square Error.

Introduction:

Fingerprints are the most popular biometric identifier. Human experts have been substituted by Automatic Fingerprint Identification Systems in fingerprint recognition and classification [22,23,24]. Fingerprint matching is a difficult problem due to large variability in different impressions of the same finger, partial overlap, non-linear distortion, variable pressure, skin condition, noise and quality of feature extraction methods, missing data of the images , specialized(forensics application) fingerprint databases. To solve this problem , we need to have a good synthetic fingerprint algorithm and also missing data reconstruction algorithm.

Recently a Gaussian Mixture model has been used in image processing applications [29] to represent a given image by combination of gaussian models. In this paper, we present a new application of Gaussian Mixture Models for Fingerprints Image Duplication and analysis. A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters.

$$s(\mathbf{x}) = \sum_{k=1}^{K} \pi_i N_k(\mathbf{x}, \mu_k, \sum_k) \text{ where}$$
$$N(x, \mu, \Sigma) = N(x, \theta) = \frac{1}{\sum_k \sqrt{2\pi}} e^{-\frac{(x-\mu_k)^2}{2\sum_k^2}} \pi_i = \frac{\pi_k N(x \mid \mu_k, \sum_k)}{\sum_{j=1}^k \pi_j N(x \mid \mu_j, \sum_j)} \text{ Where}$$

 π_i are the weights with $N_k(\mathbf{x}, \mu_k, \sum k)$ is a i-Gaussian distributions component of the mixture model with its own mean μ_k and variance shape Σ_k

The standard GMM uses a single adaptation or learning rate that is a compromise between the different rates of parameters. MMs were originally proposed in the paper by Friedman and Tee-Won Lee [1] to cope with slow-moving objects, though the standard formulation uses the update method proposed by Stauffer and Grimson [2] and [3]. In this approach each pixel is temporally modeled as a mixture of two or more Gaussians and is updated for each new image frame. The stability of the Gaussian distribution is evaluated to estimate if they are the result of a more stable background process or a short-term foreground process. Each pixel is classified to be background if the distribution that represents it is stable and above a threshold. The training of Gaussian Mixture Models can be accomplished using Expectation Maximization.

The use of a GMM for representing feature distributions in a biometric system may also be motivated by the invention that the individual component densities may model so underlying set of hidden classes. For example, in speaker recognition, it is reasonable to assume the acoustic space of spectral related features corresponding to a speaker's broad events, such as vowels, nasals or fricatives. These acoustic classes reflect some general speaker dependent configurations that are useful for characterizing speaker density.

The GMM approach has been adapted and extended by many researchers. Power and Schoonees [11] used hysteresis thresholding, introducing a faster and more logical application of the fundamental approximation of Stauffer and Grimson [10]. Other authors have improved the speed and adaptation rate of the model by updating the standard formulation . adaptively select the number of Gaussians used to model each pixel, employing a recursive computation to update the model parameters. Greggio et al. proposed a self-adaptive GMM for real-time background subtraction, learning an initial description mixture from the first video frame. Martel-Brisson and Zaccarin have extended the method to deal with shadows using a novel statistical model that copes with highly saturated scenes subject to complex, timevarying parametric, suppressing false detection in regions where shadows cannot be detected. Shah et al. [4] propsed an parametric invariant background model using GMM and SURF featutres to quickly adapt local parameters. Yoshinaga et al. [4] applied a GMM to a local difference pattern. However, these methods remain unsatisfactory for the following reasons: (i) the temporal and spatial constraints of pixel dependencies are not addressed by these pixel-based algorithms; (ii) the algorithms are computed in RGB space, but these color components are not independent and so using a simplification of the covariance by a 3×3 identity matrix is not accurate and results in more false positive and false negative detections [3]; (iii) an established model with sufficiently small variance is unnecessarily duplicated; and (iv) the proportion of time over which a pixel will observe the background is assumed constant, but in reality it fluctuates constantly depending on the number of objects and their movement patterns.

So, in order to solve this problem we need synthetic fingerprint images. Many alternatives have also been proposed to the GMM approach, including the following: an eigen background algorithm combined with a statistical parametric model; block-based oneclass background classifier ; saliency detection ; low-rank matrix factorization with iteratively re-weighted least squares (IRLS) ; self-organizing artificial neural networks and ; adaptive patchbased background modeling ; scale-invariant local ternary pattern operator and pattern kernel density estimation , to cope with parametric variation in the feature space; statistical modeling of the parametric effects ; incorporation of spatial relations of pixel pairs; color and gradient information and ; and non-parametric density estimations. However, there are also drawbacks for these methods. The assumption of model is that the foreground is only a small part of the entire image.

The goal of this article is to compare commonly used GMMs for fingerprint duplication and to develop two parametric image models for fingerprint duplication (Parametric Image Model based on Improved Adaptive, genetic Algorithm and Finite Bayesian learning, Parametric Image Model generating image statistics based on Generalized Gaussian, Finite Bayesian learning and an Image Enhancement algorithm). We need a Probabilistic Parametric Image Model because it can help in reconstruction of missing data of the images and generating synthetic images for fingerprint matching. The remainder of the paper is organized as follows. We discuss the introduction with a given GMM in Section II, and present the maximum marginal likelihood estimator based algorithms for a GMM in Section III, and for an proposed new algorithm in Section IV. Computer Simulation results are presented in Section V. Section VI concludes the paper.

State of art algorithms

Mixture models should not be confused with models for compositional data, i.e., data whose components are constrained to sum to a constant value (1, 100%, etc.). However, compositional models can be thought of as mixture models, where members of the population are sampled at random. Conversely, mixture models can be thought of as compositional models, where the total size of the population has been normalized to [1].

EM algorithm[3]:

- Form K-means clusters from a set of *n*-dimensional vectors
- 1. Set *ic* (iteration count) to 1
- 2. Choose randomly a set of K means $m_1(1), ..., m_K(1)$.
- 3. For each vector x_i , compute $W(x_i, m_k(ic)), k=1, ...K$
- and assign x_i to the cluster C_i with nearest mean.
- 4. Increment *ic* by 1, update the means to get $m_i(ic), ..., m_{\nu}(ic)$.
- 5. Repeat steps 3 and 4 until $C_{k}(ic) = C_{k}(ic+1)$ for all k

Ming-Hsuan Yang Algorithm Gaussian Mixture Models[1]

- Calculate the Inputs Observation y, joint distribution p(S,k(y, k; θ)), conditional distribution pC|Y (c|y; θ), initial values θ (0)
- Apply the EM($p(Y,S(y, c; \theta)), pS|Y(c|y; \theta)$,) algorithm.
- Choose the iteration $t \in 1, 2$
- Calculate the cost function k (t) ,S← pS|Y (s|y; θ (t−1)) (E-step)
- Calculate phase values, θ (t) ← argmaxθ ,k (t), C [pY,S(y, C; θ)] (M-step)
- Find the value of pY., If θ (t) $\approx \theta$ (t-1) then
- Calculate the final parameter θ (t)

Tee-Won Lee Generalized Gaussian Mixture Model [2]

- Each training sample is assigned to one of the clusters. Denote the assignment function by $\eta(\cdot)$. Then $\eta(i) = j$ means the ith training sample is assigned to the jth cluster.
- Find the cluster covariance matrix. arg (min(,η (X ,N (i=1 - (i)) k)
- Find the observed data (incomplete): {x1, x2, ..., xn}, where n is the sample size. Denote all the samples collectively by x.
- Complete data: {(x1, y1),(x2, y2), ...,(xn, yn)}, where yi is the cluster (component) identity of sample xi.

- Calculate the collection of parameters, θ , includes: ak , μk , Σk , k = 1, 2, ..., K. I
- Find the parameters using the likelihood function is: S(x|θ) = Xn (i=1) (log X)*K (k=1) which is the objective function of the EM algorithm

Mixture Model using Bayesian learning(MMBL)[5]

- E-Step: Calculate $S(\Phi|\Phi c)$.
- M-Step: Choose $S(\Phi|\Phi c)$.
- Find the base measure parameters: λ, λ0, observed samples: x1, ..., xn, and threshold λ[n]
- $D\{\lambda[n], s(n)\} \to \min$

The decision thresholds will be updated as follows,

- This updates the splitting threshold to a value that goes linearly with the
- initial value and the actual number of components used for the computation.
- Calculate the remaining parameters mean, variance.

Alexei A. Efros [4] Texture Synthesis

- Image Quilting can be done by calculating the value of Ej (Expectation)using the Ei,j = ei,j + min(Ei1,j1, Ei1,j, Ei1,j+1), using the dijkshtra's algorithm
- For every location, search the input texture for a set of blocks that satisfy the overlap constraints (above and left) within some error tolerance. Randomly pick one such block.
- Calculate (p) = f!0 Ireal : d(!0; !(p)) = 0

Extended Gaussian Mixture Model for Fingerprints Image Duplication

In this Section, we are going to discuss the steps that are implemented for the newly developed algorithms. The steps are as follows: Assume the image pixels as follows:

$$X = \{x_1, x_2, ..., x_N\}$$
.

The general steps for the GMM is :

Step 1: Decompose given data y[n] by combination of "similar" data by using k-mean (*k*-means clustering aims to partition the *n* observations into $k (\leq n)$ sets $\mathbf{S} = \{S_1, S_2, ..., S_k\}$ so as to minimize the within-cluster sum of squares (sum of distance functions of each point in the cluster to the K center).

Step 2: This is an iterative procedure to compute the a Probabilistic Parametric Model or Gaussian Mixture Models component($\mu, \Sigma, \Pi k$)

EM consists of two steps:

Expectation step: the new parameters are estimated using the observed data and current estimates of model parameters

Maximization step: The likelihood function is maximized under the assumption that we know the old parameters

We use the following k-mean algorithm to determine the clustered data set:

$$W(C) = \frac{1}{2} \sum_{k=1}^{K} \sum_{C(i)=k} \sum_{C(j)=k} \left\| x_i - x_j \right\|^2 = \sum_{k=1}^{K} C_k \sum_{C(i)=k} \left\| x_i - \mu_k \right\|^2$$

 $x_1,..., x_N$ are data points or vectors of observations. Each observation (vector x_i) will be assigned to one and only one cluster C(i) denotes cluster number for the *i*th observation Dissimilarity measure: Euclidean distance metric

Match the parameters that are found using the EM algorithm:

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{i=1}^n \pi_i \left(x_i - \mu_k^{new} \right) x_i \left(x_i - \mu_k^{new} \right)$$

We calculate the bias vectors as follows:

$$\alpha(biasvector1) = \frac{1}{\sqrt{2\prod_{new}\sum_{new}}} \exp[-\frac{1}{2}(x-\mu_{new})\sum_{new}^{-1}(x-\mu)]$$

$$\beta(biasvector2) = \frac{1}{\sqrt{2\prod_{new+1}\sum_{new+1}}} \exp[-\frac{1}{2}(x-\mu_{new+1})\sum_{new+1}^{-1}(x_{new}-\mu_{new})]$$

Computer simulations Images used:



Images Similarity Measure[11]

There are many image similarity measures to compare two images [11,25-28]. In this article, we use Structural Similarity Image Measure [11].

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

• Mean : $\mu_k^n = \frac{1}{N_k} \sum_{i=1}^n x_i$

• Standard deviation
$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N} (x_i - \mu_x)^2\right)^{0.5}$$

- Luminance comparison $\sigma_{xy} = \frac{(2\sigma_x \sigma_y + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)}$
- Structure comparison is conducted s(x,y) on these normalized signals $(x \mu_x)/\sigma_x$ and $(y \mu_y)/\sigma_y$

Sr. No	Algorithm	Image1	Image2
1	Original Image	O=1	
		Q I	Q=1
2	Gaussian Mixture Model		
		Q=0.3411	Q=0.3756
3	Generalized Gaussian Mixture Model (GMM)		
		Q=0.6475	Q=0.6487
4	Generalized Gaussian Mixture Model using Bayesian learning (GMMB)		
		Q=0.7692	Q=0.7641
5	Texture Synthesis	Q=0.5486	Q=0.4474

6	Genetic Algorithm	Q=0.8612	Q=0.8695
7	Improved Adaptive Algorithm	Q=0.8154	Q=0.8612
8	New Algorithm based on Generalized Gaussian and Finite Bayesian learning(GM MBF)	Q=0.8645	Q=0.8748
9	Image Enhancemen t Algorithm for Gaussian Mixture Models and Finite Bayesian Model (GMMBE)	Q=0.8745	Q=0.8687
10	New Algorithm Based on Improved Adaptive, genetic Algorithm and Finite Bayesian learning (GMMBFE)	Q=9279	Q=9249

	Algorithm	Image3	Image4
1	Original Image		

		0=1	
		V-1	Q=1
2	Gaussian Mixture Model	O=0 3445	Q=0.3718
			. Jelan.
3	GMM		
		Q=0.6437	Q=0.6472
4	GMMB	O=0 7664	
		Q=0.7004	O=0 7656
5	Texture Synthesis	Q=0.5479	
		Q 0.0 17	Q=0.4452
6	Genetic Algorithm	0-0.8671	Q=0.8645
7	Improved Adaptive Algorithm	Q=0.8173	Q=0.8629



Results (Image 1):

Sr.N	Algorithm	SSIM	Entropy	Mean
0.	C C			Square
				Error
	Original Image	1	845.74	0
1	Gaussian Mixture Models, Ming- Hsuan Yang,1999	0.3411	255.21	875.23
2	GMM, Tee-Won Lee,2005	0.6475	454.12	797.39
3	GMMB. Nicola Greggio,2010	0.7692	645.17	744.43
4	Texture Synthesis, Andrea Rau,2010	0.5496	245.12	612.48
5	Improved Adaptive Algorithm, Vahid Majidnezhad,2013	0.8692	687.32	455.78
6	Genetic Algorithm, Vahid Majidnezhad,2013	0.8397	643.21	574.53
7	GMMBF	0.8645	685.12	318.54
8	GMMBFE	0.9249	723.69	219.64

		-	-	
Sr.N	Algorithm	SSIM	Entropy	Mean
0.				Square
				Error
	Original Image	1	664.23	0
1	Gaussian Mixture	0.3756	178.74	841.23
	Models, Ming-			
	Hsuan Yang,1999			
2	GMM, Tee-Won	0.6487	369.76	752.39
	Lee,2005			
3	GMMB Nicola	0.7641	487.32	737.43
	Greggio,2010			
4	Texture Synthesis,	0.4474	278.96	656.48
	Andrea Rau,2010			
5	Improved Adaptive	0.8294	475.32	494.78
	Algorithm, Vahid			
	Majidnezhad,2013			
6	Genetic Algorithm,	0.8695	574.39	571.53
	Vahid			
	Majidnezhad,2013			
7	GMMBF	0.8675	573.12	369.54
8	GMMBFE	0.9252	627.69	245.64

Results (Image 3):

Results (Image 2):

Sr.N	Algorithm	SSIM	Entropy	Mean
0.				Square
				Error
	Original Image	1	746.39	0
1	Gaussian Mixture	0.3445	259	818.23
	Models, Ming-			
	Hsuan Yang,1999			
2	GMM, Tee-Won	0.6473	454.12	753.39
	Lee,2005			
3	GMMB. Nicola	0.7671	4617	797.43
	Greggio,2010			
4	Texture Synthesis,	0.5463	269.12	674.48
	Andrea Rau,2010			
5	Improved Adaptive	0.8664	671.32	486.78
	Algorithm, Vahid			
	Majidnezhad,2013			
	~			
6	Genetic Algorithm,	0.8356	618.21	543.53
	Vahid			
	Majidnezhad,2013			
7	GMMBF	0.8663	693.12	396.54
8	GMMBFE	0.9259	769.69	275.64

Results (Image 4):

Sr.N o.	Algorithm	SSIM	Entropy	Mean Square Error
------------	-----------	------	---------	-------------------------

	Original Image	1	963.17	0
1	Gaussian Mixture Models, Ming- Hsuan Yang, 1999	0.3411	245.21	863.23
2	GMM, Tee-Won Lee,2005	0.6475	463.12	755.39
3	GMMBNicola Greggio,2010	0.7678	663.17	771.43
4	Texture Synthesis, Andrea Rau,2010	0.5456	271.12	656.48
5	Improved Adaptive Algorithm, Vahid Majidnezhad,2013	0.8694	644.32	479.78
6	Genetic Algorithm, Vahid Majidnezhad,2013	0.8342	647.21	578.53
7	GMMBF	0.8676	697.12	394.54
8	GMMBFE	0.9296	736.69	222.64

Computational Complexity:

The softwares used for the coding pupose is MATLAB R2013a. The time complexity for the algorithm is about 3.265 seconds on a WINDOWS 7 PC which has made it relatively fast.

Conclusion:

This new algorithm presents a comparison study of Gaussian Mixture Models for Fingerprints Image Duplication and analysis. The performance of presented. Algorithms were evaluated by SSIM Index, Entropy and Mean Square Error and we obtain the best results . The state of art Gaussian Mixture Models cannot be directly apply to fingerprint image duplication problem. We get the best fit statistical model for the finger print model Images which has the minimum pixel distances as compared to the original images. The GMMBFE algorithm implemented gives the best results upto 92% of the original image for its similarity. The future work includes improving the quality

Acknowledgements:

This work is a part of a Research Project and authors are thankful to NIJ Fy 14 for funding the Project Award with Grant number 2014-IJ-CX-K003

References

- C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," IEEE Trans. on PAMI, vol. 19, no. 7, pp. 780– 785, 1997.
- 2. N. Friedman and S. Russell, "Image segmentation in video sequences: A probabilistic approach," In

Proceedings Thirteenth Conf. on Uncertainty in Artificial Intelligence, 1997.

- C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," In Proceedings CVPR, pp. 246–252, 1999.
- P. J. Withagen, K. Schutte, and F. Groen, "Likelihoodbased object tracking using color histograms and EM," In Proceedings ICIP, USA, pp. 589–592, 2002.
- S. Agaian, S. Bakhtiari and M. Jamshidi, "Local Fingerprint Image Reconstruction Based on Gabor Filtering". Proc. SPIE 8406, Mobile Multimedia/Image Processing, Security, and Applications, 10.1117/12.918384, 2012
- A. Grigoryan, S.Agaian, "Transform-based image enhancement algorithms with performance measure", Advances in Imaging and Electron Physics-130,165-242,2004
- E.A Silva, K Panetta, S. Agaian," Quantifying image similarity using measure of enhancement by entropy", Defense and Security Symposium, 65790U-65790U-12,2007
- E Wharton, S Agaian, K. Panetta, "Comparative study of logarithmic enhancement algorithms with performance measure", Electronic Imaging International Society for Optics and Photonics, 606412-606412-12,2006
- B.Silver, S.Agaian, and K..Panetta, "Contrast Entropy Based Image Enhancement and Logarithmic Transform Coefficient Histogram Shifting," Proceedings, IEEE ICASSP ,7803-8874 ,2005.
- B.Silver, S. Agaian, and Karen A. Panetta, "Logarithmic Transform Coefficient Histogram Matching with Spatial Equalization," Proceedings, SPIE Defense and Security Symposium, pp237-249, 2005.
- 11. Wang, Zhou, and Alan C. Bovik. "A universal image quality index." Signal Processing Letters, IEEE 9.3,pp 81-84,2002.
- S. Agaian, Karen Panetta, and. A.Grigoryan, "A New Measure of Image Enhancement," IASTED International Conference on Signal Processing & Communication, Marbella, Spain, 19-22, 2000.
- A. Grigoryan and S. Agaian, "Image Enhancement", Advances in Imaging and Electron Physics, Academic Press, pages 165-243, May 2004.
- A.Monnet, A.Mittal, N.Paragios and V. Ramesh, "Background Modeling and Subtraction of Dynamic Scenes", In Proceedings ICCV'03, pp.1305–1312, 2003.
- 15. Z.Zivkovic and F.van der Heijden, "Recursive Unsupervised Learning of Finite Mixture Models", IEEE Trans. on PAMI, vol.26 pp 651-656, 2004.
- Efros, Alexei A., and Thomas K. Leung. "Texture synthesis by non-parametric sampling." Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. Vol. 2. pp 1033-1038, 1999.

- 17. Elguebaly, Tarek, and Nizar Bouguila. "Bayesian learning of finite generalized Gaussian mixture models on images." Signal Processing 91.4 801-820.,2011
- Lee, Te-Won, and Michael S. Lewicki. "The generalized Gaussian mixture model using ICA." International Workshop on Independent Component Analysis (ICA'00).pp 239-244, 2000.
- Mitra, Sinjini. "Gaussian mixture models for human face recognition under illumination variations.", Applied Mathematics

Vol.3 No.12A, pp 2071-2079,2012

- Elgammal, Ahmed, David Harwood, and Larry Davis. "Non-parametric model for background subtraction." Computer Vision—ECCV 2000. Springer Berlin Heidelberg,pp 751-767, 2000
- S.Bakhtiari.,"A Novel Method for Orientation Estimation in Highly Corrupted Fingerprint Images". International Journal of Intelligent Computing in Medical Sciences & Image Processing, 6(1), 45-63.,2014
- K.Egiazarian, K. Panetta.,S.Agaian. "Image Processing: Algorithms and Systems X; and Parallel Processing for Imaging Applications II." Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Vol. 8295. 2012.
- S. Bakhtiari,S.Agaian, & Jamshidi, M. ,"Local fingerprint image reconstruction based on gabor filtering", In SPIE Defense, Security, and Sensing. International Society for Optics and Photonicspp. 840602-840602,2012..
- S.Nercessian , S.Agaian& K. Panetta.," An image similarity measure using enhanced human visual system characteristics", InSPIE Defense, Security, and Sensing . International Society for Optics and Photonics. (pp. 806310-806310),2011
- S.Nercessian , S.Agaian& K. Panetta.,"A non-reference measure for objective edge map evaluation", In Systems, Man and Cybernetics, SMC. IEEE International Conference on (pp. 4563-4568). ,2009
- C.Gao, S.Agaian& K. Panetta. "No reference color image contrast and quality measures." Consumer Electronics, IEEE Transactions on 59.3 643-651. (2013)
- 27. C.Gao, S.Agaian& K. Panetta.." Color image attribute and quality measurements.", In SPIE Sensing Technology+ Applications (pp. 91200T-91200T),2014.
- Liu, Li-min, and Tian-Shyr Dai. "Ridge Orientation Estimation and Verification Algorithm for Fingerprint Enhancement." Journal of Universal Computer Science, vol. 12, no. 10, pp 1426-1438,2006.
- 29. Yang, Ming-Hsuan, and Narendra Ahuja. "Gaussian mixture model for human skin color and its applications in image and video databases." Electronic Imaging. International Society for Optics and Photonics, pp 458-466, 1998.