An Integrated Approach for Efficient Analysis of Facial Expressions

Mehdi Ghayoumi and Arvind K. Bansal
Department of Computer Science, Kent State University, Kent, OH 44242, U.S.A.

Keywords: Emotion Recognition, Facial Expression, Image Analysis, Social Robotics.

Abstract: This paper describes a new automated facial expression analysis system that integrates Locality Sensitive Hashing (LSH) with Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to improve execution efficiency of emotion classification and continuous identification of unidentified facial expressions. Images are classified using feature-vectors on two most significant segments of face: eye segments and mouth-segment. LSH uses a family of hashing functions to map similar images in a set of collision-buckets. Taking a representative image from each cluster reduces the image space by pruning redundant similar images in the collision-buckets. The application of PCA and LDA reduces the dimension of the data-space. We describe the overall architecture and the implementation. The performance results show that the integration of LSH with PCA and LDA significantly improves computational efficiency, and improves the accuracy by reducing the frequency-bias of similar images during PCA and SVM stage. After the classification of image on database, we tag the collision-buckets with basic emotions, and apply LSH on new unidentified facial expressions to identify the emotions. This LSH based identification is suitable for fast continuous recognition of unidentified facial expressions

1 INTRODUCTION

Emotion is a psychological response to others' actions, external scene evaluations, reaction to our own thought processes possibly stirred by recall of past memories, empathy towards others emotions, or perceptions caused by variations to our sensory nerves. Emotion represents an internal state of our mind (Plutchik, 2001). Emotion profoundly affects our reactions to ourselves in near future, reactions to people we interact with, and actions to the external world consciously or unconsciously.

Emotion recognition (Crowder et. al., 2014) has become a major research area in entertainment industry to assess consumer's response and in social robotics for effective human-computer and human-robot interaction. Social robotics is an emerging area to develop new generation robots and humanoids having social acceptability. Online facial emotion recognition or detection of emotion states from video of facial expressions has applications in video games, medicine, and affective computing (Pagariya and Bartere, 2013) and (Cruz et .al. 2014).

Emotions are expressed by: (1) behavior (Plutchik, 2011); (2) spoken dialogs (Lee and Narayanan, 2005); (3) verbal actions such as variations in speech and its intensity including silence; non-verbally using gestures, facial expressions (Ekman, 1993) and tears; and their combinations. In order to completely understand the emotions, one should be able to analyze and understand the preceding events and/or predicted future events, individual expectations, personality, intentions, cultural expectations, and the intensity of an action. Many times emotions such as jealousy are hidden, and cannot be easily identified.

There are many studies to classify primary and derived emotions (Colombetti, 2009), (Gendron and Lisa, 2009), (Plutchik, 2001), and (Cambria, Livingstone and Hussain, 2012).

One popular classification of basic emotions that has been studied using facial expression analysis (Ekman, 1993) and (Pandzic and Forchheimer, 2002) identifies six primary expressed emotions: *anger, happiness, sadness, surprise, disgust* and *fear*. There are many secondary emotions such as *awe, disapproval, contempt* etc. that are derived by a combination of primary emotions (Plutchik, 2001).

Facial expressions are considered as one of the direct and fast ways of recognizing expressed emotions (Ekman and Friesen, 1978), (Fellous and Arbib, 2005), and are used between humans during

Ghayoumi M. and Bansal A..
An Integrated Approach for Efficient Analysis of Facial Expressions.
DOI: 10.5220/0005116702110219
In *Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications* (SIGMAP-2014), pages 211-219
ISBN: 978-989-758-046-8

social interactions. During verbal communications, facial expressions change continuously and are constantly inferred by the receiver (Hamm et al., 2011). Facial expressions communicate a speaker's emotions unconsciously due to the brain's reaction being translated involuntarily to facial muscles. As a consequence, interpretation of facial expressions will play an important role in the human-robot (including human-humanoids) interactions (Tian et. al., 2005).

Currently, robots do not have sufficient cognitive capabilities to analyze intentions, situations, and cultural expectations. However, they have vision and speech understanding capabilities to analyze the cues embedded in the facial expressions, dialogs, and human speech. During human-robot interactions or video display, time-efficiency is quite important to analyze facial expressions or speech analysis without losing accuracy of emotion recognition.

Automated analysis of facial expressions requires extraction of facial-features from either images or video. Two classes of techniques have been used for the analysis of facial expressions: *Facial Action Coding System* (FACS) (Ekman and Friesen, 1978) and *Facial Animation Parameters* (FAP) (Yi-Bin et al., 2006), and (Kobayasho and Hashimoto, 1995).

FACS detects the changes in the facial muscle's movements using action units (AUs) involved in facial expressions (Ekman and Friesen, 1978) and (Antonio and Indyk, 2004). FAPs techniques (Kobayasho and Hashimoto, 1995) are based upon the clues about shape and position of the features from the region of the eyes, eyebrows, forehead, lips, and mouth to recognize the presence of any of those six basic emotional states.

Techniques that consider all the facial segments suffer from increased data-space, thus making the facial expression analysis slow (Li et al., 2012). While different segments are involved in different emotions, mouth and eye-segments play a major role in analyzing the facial expressions. Although cheeks, nose and wrinkles on forehead contribute to a subset of emotions, they are strongly correlated to changes in shapes of eye and mouth segments for the corresponding emotions. In the past researchers have used dimension reduction techniques such as PCA and/or LDA (Kanade et. al., 2000), (Li et al, 2009) and heuristic methods (Yi-Bin, et al., 2006).to improve the execution efficiency of facial-expression analysis. However, none of them have reduced data-space due to redundant similar facial-images during classification and facial expression analysis.

This paper describes an automated method that improves the execution efficiency of facial expression classification and identification without loss of accuracy. Our work improves the execution efficiency by: (1) considering only mouth and eyes segments; and (2) integrating *locality sensitive hashing* with *principal component analysis* (PCA), and *linear discriminant analysis* (LDA) for removing redundant images and dimension reduction during classification of emotions.

Locality-sensitive hashing (LSH) is a method for probabilistic dimension reduction of high-dimensional data. LSH maps similar images into the same collision-buckets using a family of hashing functions, and prunes the image-space by taking only representative images from each bucket because similar images in the same bucket do not contribute much to the classification. The major contributions of the paper are:

- Pruning the image space using LSH by taking only one representative image from the cluster of similar images during the classification stage of basic emotions;
- Integration of LSH (Locality Sensitive Hashing) and K-NN (K nearest neighbours) to PCA and LDA for improving the execution efficiency during classification stage; and
- Use of LSH for identifying the emotion of untagged facial expressions.

The rest of the paper is organized as follows: Section 2 presents background about feature extraction methods and classifiers. Section 3 describes an architecture. Section 4 describes the implementation and the data-set. Section 5 describes experimental results and discussions. Section 6 discusses related works. The last section concludes the paper, and presents the future direction.

2 BACKGROUND

Facial features for analyzing facial expressions are cheeks, eyes, eyebrows, lips, mouth, nose, and wrinkles on forehead. As described in section 1, study of eye segments and mouth segments is sufficient to derive six basic emotions as described by Ekman (Pandzic and Forchheimer, 2002). Accurate localization of a facial feature plays a major role in feature extraction, and facial expression recognition.

There are many methods available for facial feature extraction, such as eye-location detection (Cevikalp, et al. 2011) and segmentation of face-area and feature detection (Dubuisson et al, 2001) (Dantone et. al., 2012). The major segments that express all six basic emotions are: (1) two eye-segments including eyebrows; and (2) mouth segment including lips.

2.1 Locality Sensitive Hashing (LSH)

Locality-sensitive hashing (LSH) is a method for probabilistic dimension reduction of high-dimensional data. In LSH, similar input items are mapped to the same collision-bucket using a family of locality-sensitive hash functions (Andoni and Indyk, 2008). Consider a family $\mathcal{H}$ of hash functions mapping $\mathbb{R}^d$ (d-dimensional space) to some universe U. A point p is an *R-near* neighbour of another point q if the distance between p and q is at most *R*. According to LSH, a family $\mathcal{H}$ is called (R, cR(c is approximation factor), P_1, P_2) sensitive if for any p, q ∈ $\mathbb{R}^d$, following two conditions hold.

$$\text{If} \|p - q\| \leq R \text{ then } Prob_H[h(q) = h(p)] \geq P_1 \quad (1)$$

$$\text{If} \|p - q\| \geq cR \text{ then } \mathrm{Prob_H}[h(q) = h(p)] \leq P_2 \quad (2)$$

Where h ∈ $\mathcal{H}$ is a hash function; *R* is the cut-off distance for nearness; P_1 and P_2 are probability values that lie between 0…1. Equations (1) and (2) state that if the two data-points (p and q) lie within the cut-off-distance R then the probability of similarity is high, and points p and q will map on the same collision-bucket. LSH is beneficial if P_1 is greater than P_2.

2.2 Principle Component Analysis (PCA)

Principal component analysis is a dimension reduction technique based upon the transformation of coordinate-axis along the direction of the data-points with maximum variance. This transformed coordinate-axis becomes the first principal component. The second principal component is orthogonal to the first principal component with the next maximum variance. This way only a few major components are needed for data-point analysis.

For image analysis, image samples are trained to obtain the principal subspace composed of orthogonal basis vectors, then mapping the samples into the subspace to derive projection coefficient vectors as sample feature-vectors. Test images are mapped into the principal subspace to derive the corresponding feature-vectors of the test-images.

A face is modelled as a pixel-matrix X. The Eigen-decomposition on XX^T derives two matrices $U\ and\ V$ such that U is an n × n matrix whose columns are the eigenvectors of XX^T, and V is an N × N matrix whose columns are the eigenvectors of X^TX. An r-dimensional subspace is formed by selecting the first r rows of the transformed data-matrix X_{LD} as follow

$$X_{LD} = U^tX \quad (3)$$

The N × N covariance matrix XX^t gives:

$$X^tX = V\Lambda^{1/2}U^tU\Lambda^{1/2}V^t = V\Lambda V^t \quad (4)$$

Where Λ is the covariance matrix of XX^T and:

$$X_{LD} = XV \quad (5)$$

2.3 Linear Discriminant Analysis (LDA)

LDA is a dimension reduction technique that preserves discriminatory information such as scatter-matrices and their mixture. Scatter-matrix is an estimation of covariance matrix for multivariate normal distribution. LDA is used for two or more classes of objects. LDA aims at increasing separability of the samples in the subspace. LDA criteria are mainly based on a family of functions of scatter-matrices.

LDA uses two types of scatter-matrices: (1) within-class scatter matrix denoted by Σ_b, and (2) between-class scatter matrices denoted by Σ_w. LDA searches for a group of basis vectors, which make different class samples, and have the smallest within-class scatter and the largest between-class scatter. The equations for the two scatter-matrices are

$$\Sigma_b = \sum_{i=1}^{g} N_i(\bar{x_i} - \bar{x})(\bar{x_i} - \bar{x})^T \quad (6)$$

$$\Sigma_w = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (x_{i,j} - \bar{x_i})(x_{i,j} - \bar{x_i})^T \quad (7)$$

Where $x_{i,j}$ is a data point in i^{th} class, $\bar{x_i}$ is the mean of each class, $\bar{x}$ is the overall mean, N_i are the data points in each class, and *g* is the number of classes. The main objective of LDA is to find a projection matrix that maximizes the ratio of the determinant of Σ_b to the determinant of Σ_w.

Alternatively, $\Sigma_w^{-1}\Sigma_m$ can be used for LDA, where Σ_m represents the mixture scatter-matrix ($\Sigma_m = \Sigma_b + \Sigma_w$). The computation of the eigenvector matrix ϕ from $\Sigma_w^{-1}\Sigma_m$ is equivalent to the solution of the generalized eigenvalue problem:

$$\Sigma_m \phi = \Sigma_w \phi \Lambda \tag{8}$$

Where Λ is an eigenvalue matrix. Final transformation is achieved by:

$$y_i = \phi x_i \tag{9}$$

The low-dimension vector y_i is the LDA feature of the sample x_i (Cristianini and Shawe-Taylor, 2000).

3 OVERALL ARCHITECTURE

Our classification system (see Figure 1) has four stages: 1) eyes and mouth-segment detection; 2) clustering using LSH to prune redundant similar images; 3) feature extraction using PCA and LDA to reduce the dimensions; and 4) classification by training SVM and K-NN. The input to the face-segment stage is an image database with tagged emotion for classification. The database of eye-segments and mouth-segment forms the input for the LSH-stage. The LSH-stage produces a set of non-redundant images that becomes the input for the dimension-reduction stage. Dimension-reduction stage generates transformed Eigen-images (see Figure 7a and 7b). These Eigen-images are used to train the SVM for the classification of emotions. After the classification, collision-buckets are labelled with one of the primary emotions, and LSH is used on online images to identify the untagged facial expressions.

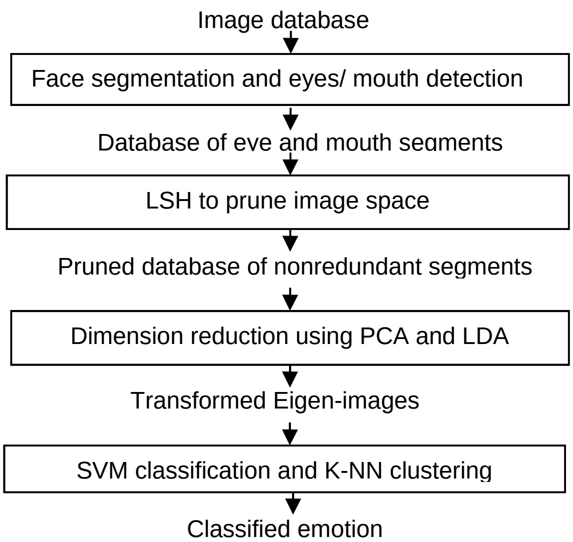


Figure 1: An architecture for the model.

3.1 Detection of Eye and Mouth Segment

During this stage, the software compares different features such as grey-level values, distance transform features, luminance, color and edge properties for automatic localization of eyes and mouth (Cevikalp et. al., 2011), (Beigzadeh and Vafadoost, 2008), or (Jianke et. al., 2009). We have used Cevikalp's method (Cevikalp, et al., 2011) to find eye and mouth segments in a face. This algorithm automatically extracts geometric and texture features using grey-level values and distance transform features.

3.2 Application of LSH

During stage 2, LSH is applied to derive nearest neighbour clusters on the database of *mouth-segment* and *two eye-segments* (see Figure 2).

LSH makes multiple buckets each containing similar images. Similarity is measured by hamming distance of the hash-key of the colliding images. Figures 3 and 4 show a set of similar images for eye-segments and mouth segment in the same bucket. Due to similarity of images, we just use one image from each bucket; other images are pruned since similar images do not contribute additional information during classification. Pruning redundant similar images improves the computational efficiency significantly as shown in Figure 8 in Section 5.

Figure 5 shows mapping of the images in the database to the collision-buckets after LSH. As shown in the figure, the number of images per bucket increases significantly after 450 images showing greater collision as the number of images increase.

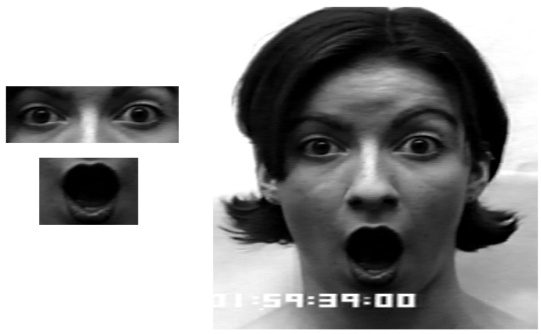


Figure 2: Original image and three part for training.

After LSH processing, M images in the database are mapped on m (m << M) collision-buckets. Each collision-bucket contains similar images. Only one representative image is taken from each bucket. This set of representative images from each bucket becomes the input for LDA and PCA stage. Thus the number of images processed in PCA and LDA stage reduces to m images.

LSH employs an approximate similarity strategy that reduces the images by: (1) putting similar images into the same collision-buckets; and (2) picking only one image from a bucket. M images from the database are reduced to m (m << M) images from different collision-buckets. Under the assumption that LSH uniformly distributes at least one image in a bucket, our scheme removes the frequency-bias during computation of principal components during PCA caused by non-uniform distribution of images in the image database.

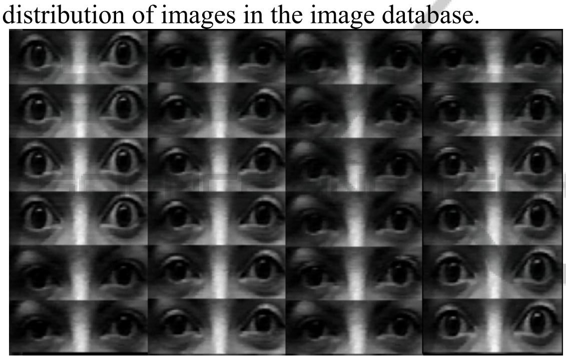


Figure 3: Eye-images in a collision-bucket.

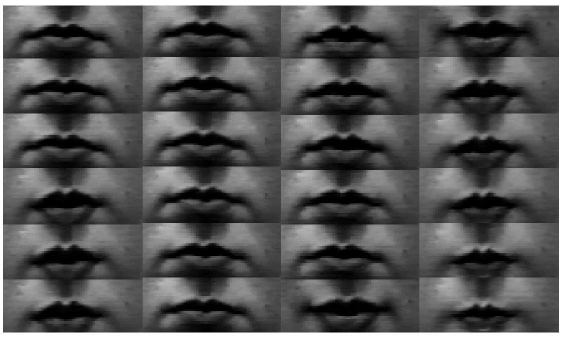


Figure 4: Mouth-images in a collision-bucket.

3.3 Dimension Reduction

Mapping the data-space to feature-vectors reduces redundant information while preserving most of the intrinsic information content of the pattern (Abrishami and Ghayoumi, 2006), (Kanade et. al., 2000). Figures 6a and 6b show PCA Eigen-images for mouth and eyes segments. Figures 7a and 7b illustrate 16 images transformed by LDA for training our classifiers.

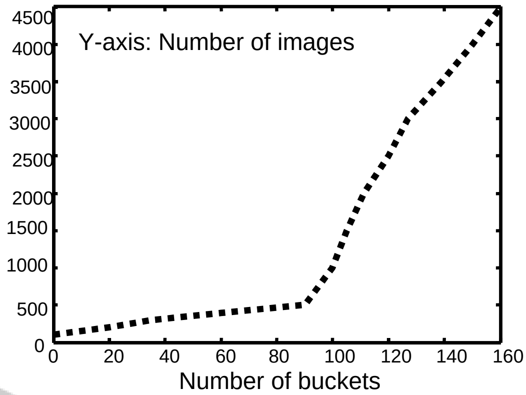


Figure 5: LSH performance (images vs. buckets).

3.4 Classification using SVM and KNN

The recognition module consists of two steps: (i) features extraction from test data and (ii) classification. We use Support Vector Machine (SVM) to classify test-data into emotion classes. Reformulating the problem using the Lagrangian, the expression to optimize for a nonlinear SVM (Cristianini and Shawe-Taylor, 2000) is as follows:

$$L(\alpha) = \sum_{i=1}^{r} \alpha_i - 1/2 \sum_{i=1}^{r} \sum_{j=1}^{r} \alpha_i \, \alpha_j y_i y_j K(x_i, x_j) \qquad (10)$$

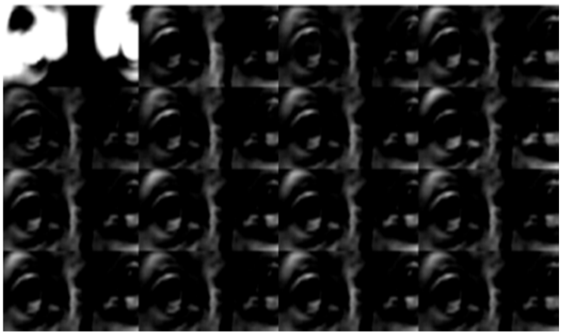
Where α_i is a Lagrangian multiplier, $y_{i,}\ y_j \in \{-1,\ 1\}$, and $x_i,\ x_j \in R^n$ are points in the data-space. $K(x, x')$ is a kernel-function satisfying Mercer's conditions. An example of kernel-function is the Gaussian radial basis function:

$$K(x, x') = exp\left(\frac{\|x - x'\|^2}{2\sigma^2}\right) \qquad (11)$$

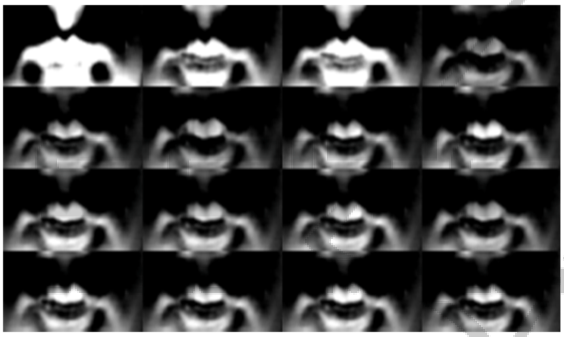
Where σ is the standard deviation. The decision function of the SVM is described by:

$$f(x) = sgn\left[\sum_{i=1}^{p} y_i \alpha_i \, K(x, x_i) + b\right] \qquad (12)$$

Where sgn (signum) function extracts the sign of a real number, b is the bias, and α_i denotes a support vector. For data-points that lie closest to the optimal hyperplane, the values of corresponding support vector α_i is non-zero. We used the K-Nearest Neighbour algorithm (KNN) to classify a test-data into emotion-classes. Given an instance of a test-data X, KNN derives k neighbours nearest to the unlabelled data from the training set based on a distance measure, and labels the new instance by looking at its nearest neighbours. In our case, the distance measure is Euclidean distance.

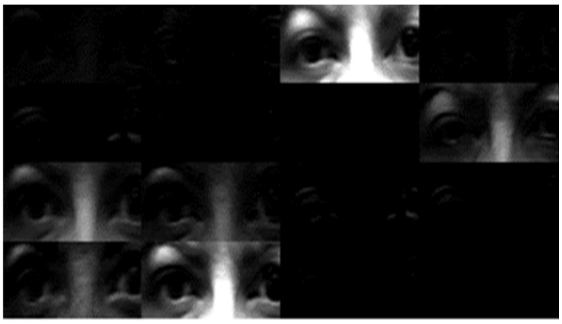


(a)



(b)

Figure 6: PCA features (a) eyes, (b) mouth.



(a)

(b)

Figure 7: LDA features (a) Eyes, (b) Mouth.

3.5 Tagging Unidentified Expressions

After the classification, the collision-buckets get tagged with specific basic emotions. An unidentified facial expression is tagged with a basic emotion by applying LSH. The untagged facial expression maps into one of the collision-buckets, and the emotion tag of the collision-bucket becomes the emotion tag of the unidentified facial expression.

4 IMPLEMENTATION

4500 facial-expression images have been taken from the CK+ database (Lee and Narayanan, 2005). Each image has 50*170 pixels, and the images are already tagged with the basic emotions. For each emotion category, 1/3rd of the images have been selected for training, and the remaining images have been used for testing. The images have been reshaped into 8500*1 dimensional column vectors. E^2LSH software package (Andoni and Indyk, 2004) has been used for LSH hashing. The LSH reduces 4500 database images to 164 images. The average reduction ratio for the images is 26. These images are fed for feature extraction and dimension-reduction in features-space using PCA and LDA. MATLAB has been used for developing related PCA and LDA software in our experiments. LIBSVM library (Chang and Lin. 2001) has been used for multiclass strategy. Finally, the results derived by our approach, naïve SVM approach and naïve KNN approach have been compared with the tagged emotions in the CK+ database to derive the percentage accuracy.

5 RESULTS AND DISCUSSION

Tables 1, 2, and 3 show the correctness of emotion recognition by our integrated approach, by applying just Support Vector Machine (SVM) technique, and by applying just K-Nearest Neighbours (KNN) technique, respectively. Our integrated approach shows better accuracy compared to naive KNN or SVM based recognition due to the removal of the effect of redundant similar images at LSH stage that led to removal of frequency-bias caused by non-uniform distribution of similar images in the database during PCA and SVM classification stages. Figure 8 shows the time efficiencies of "KNN with PCA + LDA", "SVM with PCA + LDA", "KNN with LSH + PCA + LDA", and "SVM with LSH + PCA + LDA". The use of LSH reduces

Table 1: Correct emotions using integrated approach.

State	PCA+LDA	State	PCA+LDA
Happiness	83%	Surprise	95%
Sad	90%	Anger	84%
Fear	84%	Disgust	93%

Table 2: Correct emotions using just SVM.

State	PCA+LDA	State	PCA+LDA
Happiness	83%	Surprise	85%
Sad	83%	Anger	78%
Fear	78%	Disgust	80%

Table 3: Correct emotions using just KNN.

State	PCA+LDA	State	PCA+LDA
Happiness	77%	Surprise	76%
Sad	77%	Anger	72%
Fear	73%	Disgust	73%

the image space by a factor of 26; improving the performance by more than an order of magnitude.

6 RELATED WORKS

There are three approaches for facial feature extractions: (1) geometric feature-based methods, (2) appearance-based methods (Tian et. al., 2005), and (3) Gabor filters (Dahamane and Meunier, 2011). Many researchers have used dimension reduction techniques (either PCA or LDA) to reduce the data- space (Kanade et. al., 2000), (Yi-Bin, et al., 2006), or a combined approach (Li et al., 2012). However, they do not use LSH to reduce images and identify emotions. LSH was first proposed by Indyk and Motwani (Indyk and Motwani, 1998) for fast approximate search. Vision researchers have shown the effectiveness of LSH based fast retrieval for image search applications including shape matching and pose inference (Shakhnarovich, Darrell and Indyk, 2006) (Jain, Kulis and Grauman, 2008). We integrate LSH with PCA and LDA to improve the execution efficiency without loss of accuracy.

For facial expression recognition, existing methods either aim at developing feature selection techniques or designing novel classification algorithms for improved performance (Sun et. al, 2009). Dahmane and Meunier proposed an approach for representation of the response to a bank of Gabor energy filters with histograms, and applied SVM with a radial basis function for classification (Dahmane and Meunier, 2011). Other recent works (Li et. al, 2012) identify representative regions of face images.

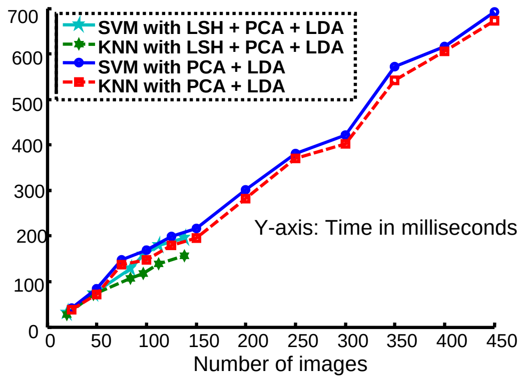


Figure 8: Time-efficiencies in difference schemes.

Cruz et. el. (Cruz et. el. 2014) are working on applying attention theory to improve accuracy of emotion recognition. Our work is based upon data-space reduction, and complements their work. Both work can be combined to benefit from each other.

In our present work, we have used LSH to prune redundant similar images that do not contribute to image classification. This improves the execution efficiency of the emotion detection as well as the accuracy of the emotion detection as illustrated in Tables 1- 3 by removing the frequency-bias of the redundant similar images during dimension reduction stage in PCA and SVM classification stage. We will benefit from the works by (Li et. al, 2012) by providing dynamic weights to the facial-expression analysis especially for online video analysis where multiple consecutive frames show very similar emotions.

7 CONCLUSION

We have implemented an integrated approach for recognizing Ekman's basic emotion categories using facial expression images. By focusing on major segments needed for facial expressions analysis and by reducing the number of redundant similar images using LSH, we derive a non-redundant set of images for training. The use of representative images from LSH buckets reduces the effect of frequency bias during PCA stage and SVM classification stage. Our experimental results confirm that LSH based approach improves the execution efficiency of emotion classification and recognition without loss of accuracy.

Currently, our LSH based scheme uses hard clustering by mapping facial expressions on basic emotions. The scheme needs to be extended to handle secondary emotions that are mixtures of primary emotions. We also need to extend our scheme to learn new mixed emotions by video analysis and online analysis. We are currently extending our techniques to handle secondary and mixed emotions. We are also looking at removing redundancy in consecutive frames in video analysis.

REFERENCES

Abrishami, M. H., Ghayoumi, M., 2006. “Facial Image Feature Extraction using Support Vector Machine,” International Conference of Vision Theory and Applications, Lisbon, Portugal, pp. 480 - 485.

Andoni, A., Indyk, P., 2008. “Near Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions,” Communications of the ACM, Vol. 51, No. 1, pp. 117 - 122.

Andoni, A., Indyk, P., 2004. "E2lsh: Exact Euclidean locality-sensitive hashing," implementation available at http://web.mit.edu/andoni/www/LSH/index.html.

Beigzadeh, M., Vafadoost, M., 2008. "Detection of Face and Facial Features in digital Images and Video Frames," Proc. of the Cairo International Biomedical Engineering Conf. (CIBEC'08), Cairo, Egypt.

Cambria; E., Livingstone, A., Hussain A., 2012. "The Hourglass of Emotions". Cognitive Behavioural Systems. Springer, pp. 144 - 157.

Cevikalp, H., Yavuz, H. S., Edizkan, R., Gunduz, H., Kandemir, C. M., 2011. “Comparisons of features for automatic eye and mouth localization,” Thirty First SGAI International Conference on Innovations in Intelligent Systems and Applications (INISTA), Cambridge, UK, pp. 576 - 580.

Chang, C. C., Lin, J., 2001, “LIBSVM: a Library for Support Vector Machines”, http://www.csie.ntu. edu.tw/ ~cjlin/libsvm/.

Colombetti, G., 2009. "From affect programs to dynamical discrete emotions". Philosophical Psychology 22: 407-425.

Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines, Cambridge University Press, UK.

Crowder, J. A., Carbone, J. N., Friess, S. A., 2014. Artificial Cognition Architectures, Springer Science + Business Media, New York.

Cruz, A., Bhanu, B., Thakoor, N., 2012. "Facial emotion recognition in continuous video," Proc. of the Int’l. Conf. Pattern Recognition, Tsukuba, Ibaraki, Japan, pp. 2625-2628

Cruz, A., Bhanu, B., Thakoor, N., 2014. "Vision and Attention Theory Based Sampling for Continuous Facial Emotion Recognition," IEEE Transactions of Affective Computing, to appear, pre-print available at http://www.ee.ucr.edu/~acruz/j1-preprint.pdf

Dahmane, M. and Meunier, J., 2011. “Continuous emotion recognition using Gabor energy filters,” ACII'11 Proc. of the 4th international conference on Affective computing and intelligent interaction - Volume Part II, Springer-Verlag, Berlin / Heidelberg, pp. 351 - 358.

Dantone, M., Gall, J., Fanelli, G, Gool, L. V., 2012. "Real Time Facial Feature Detection using Conditional Regression Forests, IEEE Conf. on Computer Vision and Pattern Recognition, RI, USA, pp. 2578-2585.

Dubuisson, S., Davoine, F., Cocquerez, J., 2001. "Automatic Facial Feature Extraction and Facial Expression Recognition," 3rd International Conf. of Audio and Video-Based Biometric Person Authentication, Halmstad, Sweden, Springer, pp. 121-126.

Ekman, P., Friesen, W.V., 1978. The Facial Action Coding System, Consulting Psychologists Press, Palo Alto, CA. Ekman, P., 1993, “Facial expression and emotion”, American Psychologist, Vol. 48, Issue 4, pp. 384 - 392. Fellous, J. M., and Arbib, M. A., 2005. Who Needs Emotions? The Brain Meets the Robot, Oxford University press, 2005.

Gendron, M., Lisa, F., 2009. "Reconstructing the Past: A Century of Ideas about Emotion in Psychology," Emotion Review 1 (4): 316-339.

Hamm, J., Kohler, C. G., Gur, R. C., Verma, R., 2011. “Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders,” Journal of Neuroscience Methods, Vol. 200, Issue 2, pp. 237 - 256.

Jain, P., Kulis, B., Grauman, K., 2008. Fast Image Search for Learned Metrics. In CVPR.

Jianke, L., Baojun, Z., Zhang, H., Jiao, J., 2009. ”Face Recognition System Using SVM Classifier and Feature Extraction by PCA and LDA Combination,” International Conference on Digital Object, San Francisco, CA, USA, pp.1 - 4.

Kanade, T., Cohn, J., Tian, Y., 2000. “Comprehensive database for facial expression analysis,” Fifth IEEE International Conf. on Automatic Face and Gesture Recognition, Grenoble, France, pp 46 - 53.

Kobayasho, S., Hashimoto, S., 1995. “Automated feature extraction of face image and its applications,” International workshop on Robot and Human Communication, Tokyo, Japan, pp. 164 - 169.

Lee, C. M., Narayanan, S. S., 2005. "Towards Detecting Emotions in Spoken Dialog," IEEE Trans. on Speech and Audio Processing, Vol. 13, No. 2, pp. 293-303.

Li, J., Zhao, B., Zhang, H., Jiao J., 2009. "Dual-Space Skin-Color Cue Based Face Detection for Eye Location," International Conference on Information Engineering and Computer Science (ICIECS 2009), pp.1 - 4.

Li, J., Zhao, B., Zhang, H., Jiao J., 2009. "Face Recognition System Using SVM Classifier and Feature Extraction by PCA and LDA Combination,"

International Conf. on Computational Intelligence and Software Engineering (CiSE 2009), pp. 1 - 4.
Li, R., Hu, M., Wang, X., Xu, L., Huang, Z., Chen, X., 2012. “Adaptive Facial Expression Recognition Based on a Weighted Component and Global Features,” Fourth International Conf. on Digital Object, Guanzhou, China, pp. 69 - 73.
Pagariya, R. R., Bartere, M. M, 2013. "Facial Emotion Recognition in Videos Using HMM," International Journal of Computational Engineering Research, Vol, 03, Issue, 4, Version 3, pp. 111-118.
Pandzic, I. S., Forchheimer, R., 2002. MPEG-4 Face Animation, the Standard, Implementation and Applications, Wiley, pp. 17-55.
Perakyla, A., Ruusuvuori, J., 2012. “Facial Expression and Interactional Regulation of Emotion,” Emotion in Interaction, Oxford University Press, pp. 64 - 91.
Plutchik, R., 2001. "The Nature of Emotions". American Scientist, Vol. 89, pp. 344 - 350.
Shakhnarovich, G., Darrell, T., Indyk, P., editors. Nearest-Neighbor Methods in Learning and Vision: Theory and Practice. The MIT Press, 2006.
Tian, Y. L., Kanade, T., Cohn, J. F., 2005. Facial Expression Analysis, Handbook of Face Recognition, pp. 247-275.
Yi-Bin, S., Jian-Ming, Z., Jian-Hua, T., Geng-Tao, Z., 2006. “An Improved facial feature localization method based on ASM," Proc. of the 7th International Conference on Computer Aided Industrial design and Conceptual design, Hangzhou, China, pp. 1-5.