Random Sampling for Patch-based Face Recognition

Ismahane Cheheb*, Noor Al-Maadeed†, Somaya Al-Madeed†, Ahmed Bouridane *, Richard Jiang*
* Computer and Information Science, Northumbria University. Newcastle Upon Yune, UK.
{ ismahane.cheheb , ahmed.bouridane , richard.jiang }@northumbria.ac.uk
† Dept. of Computer Science and Engineering, Qatar University. Doha, Qatar.
{n.alali, s_alali}@qu.edu.qa.

***Abstract*—Real face recognition is a challenging problem especially when face images are subject to distortions. This paper presents an approach to tackle partial occlusion distortions present in real face recognition using a single training sample per person. First, original images are partitioned into multiple blocks and Local Binary Patterns are applied as a local descriptor on each block separately. Then, a dimensionality reduction of the resulting descriptors is carried out using Kernel Principle Component Analysis. Once done, a random sampling method is used to select patches at random and hence build several sub-SVM classifiers. Finally, the results from each sub-classifier are combined in order to increase the recognition performance. To demonstrate the usefulness of the approach, experiments were carried on the AR Face Database and obtained results have shown the effectiveness of our technique.**

I. Introduction

Face recognition is a hot research topic in computer vision and biometric security. Out of all biometric modalities, face is the most popular as it can be captured easily at a distance and has numerous applications in security, surveillance, access control...etc. Various techniques have been proposed for effective face recognition under various distortions. Conventional methods use holistic features of the face where the information extracted from the whole face is taken into account using global linear and non-linear statistical techniques. Principal Component analysis (PCA) method [1], Independent Component Analysis (ICA) [2], and Linear Discriminant Analysis (LDA) [3] are the most popular linear techniques proposed by the researcher community. However, these methods are not efficient due to the non-linear characteristic of the face. Non-linear kernel-based techniques, namely Kernel PCA (KPCA) and Kernel Fisher Analysis (KFA) [4], [5] have been proposed to solve the problem by taking full advantage of the face's contour and the curve detail information.

More recent methods use local descriptors [6] which, in contrast to global descriptors, represent the features in local regions and have proved to be more effective. Another method is the patch-based face recognition, first presented in [7], which works by partitioning an image into multiple overlapping or non-overlapping patches before applying global or local descriptors for use in the the recognition stage.

In patch-based approaches, local features are extracted from each region, called patch, of the image. The face images can be divided into either overlapping or non-overlapping blocks.

A number of approaches for patch-based face recognition have been proposed previously. For example, the work in [8] proposes a feature concatenation and block selection with similarity measures while the authors in [9] propose to use the weight of the classification results of the patches by computing the correct classification rates on the test set. [10] employs subspace methods and uses a majority voting to combine the classification results of the patches and random subspaces. The authors in [11] train the classifiers from separate patches before combining them by a two-step layer decision maker: the first one is a weighted summation while the second is a combination of the local ensemble classifiers decision with a global classifier obtained from the whole face. The study described in [12] selects face areas that contain more information compared to other areas. Even though this approach is very effective and has proved to be more robust to the challenges of illumination variations and partial occlusions, it is still lacking in terms of classification performances since one single classifier is constructed for all the image patches. In [13], the authors find the largest matching area at each point of the face. However, in order to work well with partial occlusions, this method goes through an occlusion de-emphasis step.

To solve this problem, we propose the use of Random Patch Sampling, using all face patches equally by building multiple classifiers aiming to improve the classification accuracy.

This paper presents a patch-based face recognition method. First, we partition the image into a number of 50%-overlapping regions. Next, we use LBP as a feature descriptor independently on the image patches. To reduce the resulting high dimensionality of the descriptors, KPCA is employed. Once done, the feature vectors of the image patches are normalised before we randomly sub-sample a number of blocks within each image and build multiple SVM classifiers. To validate the approach, experiments were carried using a single sample per person (SSPP) as per real world conditions. The final results are obtained by combining the performances from all the sub-classifiers with a union rule. fig.1 illustrates the process.

The remainder of this paper is presented as follows: Section II introduces the face patching concept. Section III describes the Multi-LBP method and its application to our work. Section IV presents the Kernel PCA method while Section V discusses the proposed Random Patching method and its use for the application at hand. Experiments carried out are discussed in

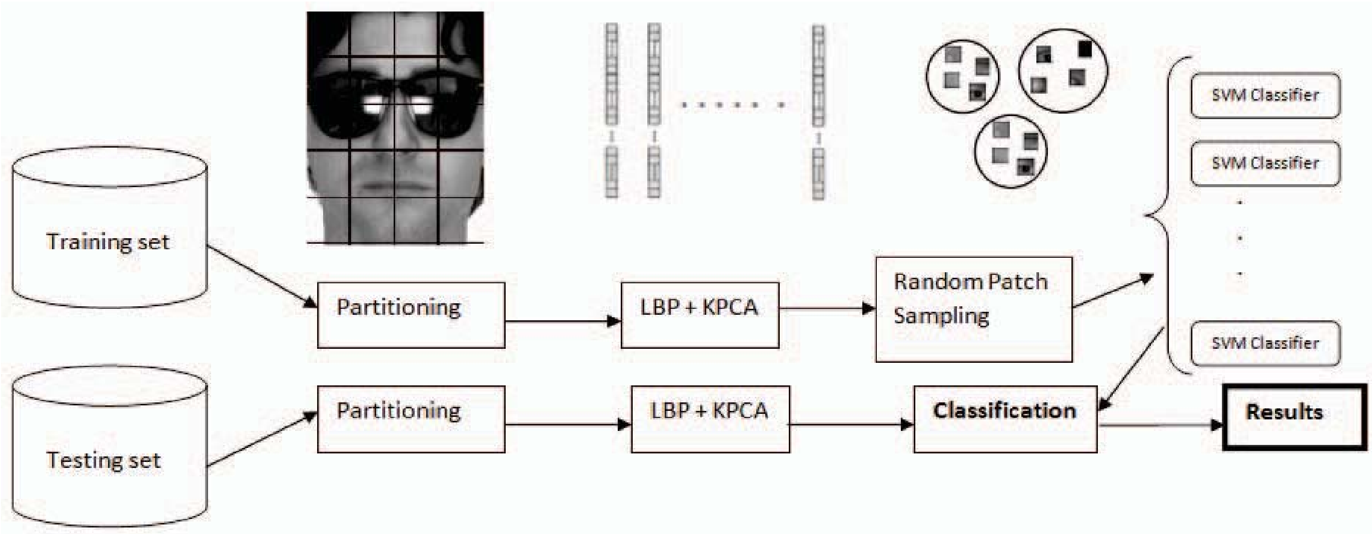

Fig. 1. Face Recognition Based on Random Patch SVM

Section VI. Finally, conclusion and future work are in Section VII.

II. Face Patching

In our work a face image is defined as a set of multiple overlapping and non-overlapping blocks. In order to select the appropriate block size an important step that can affect the recognition significantly, initial experiments were conducted by varying the block size. It is to be noted that the blocks need to provide enough discriminant information about the region but also they need to prevent confusion at the feature extraction . To achieve these constraints, our initial analysis has shown that 33x30 overlapping patches provide the best results.

III. Multi-scale Local Binary Patterns

The LBP operator is a widely used local texture descriptor whether in face recognition [14] or other applications. It is a general combination of gray-scale invariants and operates by assigning a label to every pixel of an image by thresholding the (P,R) neighbourhood (P sampling points on a circle of radius R) of each pixel with the centre pixel value and considering the result as a binary number. The obtained histogram of the labels is used as a texture descriptor. The first neighbourhood introduced in LBP was P=8 and R=1 as in thresholding the 8 neighbouring pixels in a radius of 1 as shown in fig.2 This was later extended to use neighbourhoods of different sizes.

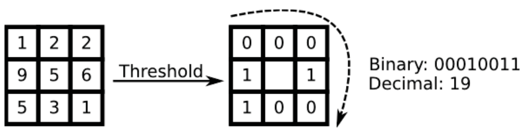

Fig. 2. LBP Operator

As illustrated by fig.2, the threshold can be obtained by comparing the neighbourhood pixel gp with the centre pixel gc. The operator gives 1 if the gp is larger than gc and 0 otherwise. The final form of the LBP is in decimal value. The features extracted by the LBP operator are represented in a histogram. This process could be mathematically expressed as

$$LBP_{P,R} = \sum_{p=0}^{P-1} f(g_p - g_c)2^P, f(x) = \begin{cases} 1, & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (1)$$

LBP histograms for face description have been presented in [15] where the face is divided into multiple local regions and the texture descriptors are extracted from each area and which are combined into one uniform histogram representing the face image as described in fig.3 Uniform histograms were inspired from the fact that some binary patterns occur more commonly in facial image than others and are therefore used to reduce the usual length of 256-bins patterns to 59 patterns [16].

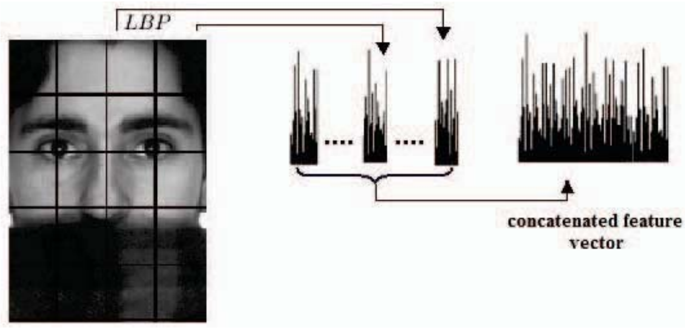

Fig. 3. LBP on a face image

However, since the area covered by the classic algorithm is usually small, we have opted to use a uniform multi-scale LBP. This means that we use different neighbourhoods: sample points $P = 8$, $P = 16$ and $P = 24$ with different radii ranging from $R = 1$ to $R = 12$ and which are later on concatenated into one uniform LBP histogram of over 4000 bins in order to cover a larger area and eventually provide a larger range of discriminative features. Containing a certain number of non-informative features is not an issue given the fact that we are reducing dimensionality in our next step.

IV. Kernel Principal Component Analysis

Kernel principal component analysis (KPCA) is a non-linear feature extraction technique. It is an extension of principal component analysis (PCA) using kernel methods. The non-linearity is presented through the mapping of the data from the input space to a feature space F.

$$\Phi : R^N \rightarrow F \tag{2}$$

It is, therefore, possible to map an input variable into a high-dimensional feature space before performing PCA. Applying the PCA in a high-dimensional feature space will result in high-order statistics of the input variables. It is, however, difficult and computationally intensive to compute the dot products of vectors in the high-dimensional feature space. As a solution, kernel methods have been used computing the dot products in the original low-dimensional input space using a kernel function therefore overcoming this difficulty.

KPCA has been widely used in the field of face recognition [17], [18], [19] especially for expression and illumination variations.

In our work, we have used KPCA with the polynomial kernel as it has been proven to work effectively when extracting facial features.

V. Random Patch-Based SMV

In previous works based on patched faces, the authors either use all the patches or select a certain number of blocks to build a global classifier. In our approach, we have chosen to use a random sampling method to build more than one classifier. A random sample is a subset of a population selected so that all samples have equal occurrence probability.

Support Vector Machines (SVMs) [20], which have been used in our classification stage, have been found to be very effective. It is a supervised learning algorithm where the system is trained by mapping the training set feature vectors in a space separating them clearly using kernel functions (Polynomial, Gaussian...). The testing set is mapped on to the same space. An SVM predicts the possible class using the in-between maximum distance.

The idea is then to train multiple SVM classifiers based on the sub-training sets and combine the individual results with a union rule to obtain the final score.

VI. Experiments and Analysis

To assess the effectiveness of our proposed approach, experiments were performed using the cropped AR face database [21], which contains 2600 images of 100 individuals (26 different images per person) under various facial expression, lighting and occlusion conditions, taken in 2 sessions. The images were then resized into 165x160 pixel in this experiment. A sample of the images in the AR database are in fig.4.
rating

Fig. 4. Sample of the AR Face Dataset

For the training stage, we have used a single clean image per person from the first session. Sunglasses and scarf occluded faces from both sessions are used for testing. See fig.5

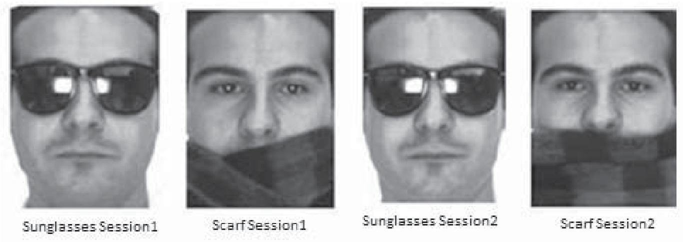

Fig. 5. Examples of faces used for testing from the AR dataset

The first set of experiments was carried out using non-overlapping patches and alternating the number of blocks per image. In fig.6, the horizontal axis indicates the number of patches used and the vertical axis represents the correct face recognition rate. We have tried the following sizes : 55x40, 33x30, 33x20 and 15x15. From the results obtained it can be seen that as the blocks become smaller and consequently their number larger, the recognition rate either drops or stabilises. It can also be noticed that when the block size is large, the recognition rate is rather low. This can be explained by the fact that large blocks contain too many features that may induce confusion at the recognition stage unlike in the case of smaller blocks.

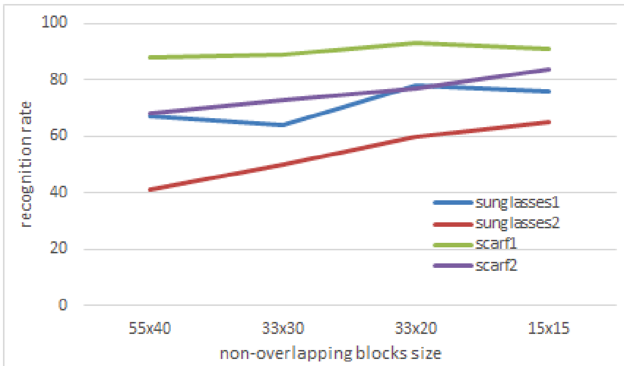

Fig. 6. Recognition performance using different patch sizes

The second set of experiments was based on the image size with non-overlapping patches. The results are depicted in fig.7 where it can be seen that the use of an original size of 165x120 gives superior performances to when using other smaller sizes.

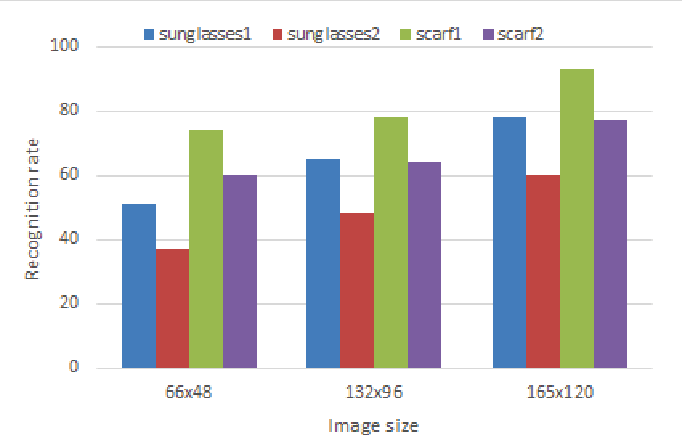

Fig. 7. Recognition performance using different image sizes

For the third set of experiments, the images are divided into 50% overlapping uniform blocks of size 33x30 pixels varying the number of features as shown in fig.8. From the results it can be seen that the recognition rate increases steadily as the number of features increases before reaching a stable rate. It can also be noticed that when the number of features is too small (under 100), the recognition rate is below 50% for all test samples.

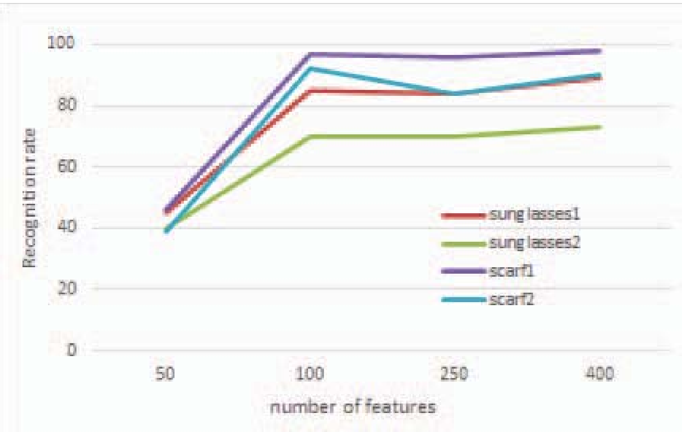

Fig. 8. Recognition performance using different numbers of features

In the fourth set of experiments, a different number of patches are used to train the classifiers: from each training face, a specific number of blocks is randomly selected to build the classifiers. Therefore, each patch from each image has an equal chance of being used. The results obtained are depicted in fig 9 showing that when using 12 patches per classifier , the recognition rate gives 86% when the scarf occludes the faces compared to a 95% when using 8 patches per classifier. On the other hand, using 4 patches per image increases the recognition rate to an outstanding 98%. This can be explained by the fact that the use of fewer patches helps reduce the confusion at the classification phase thus improving the recognition rate significantly.

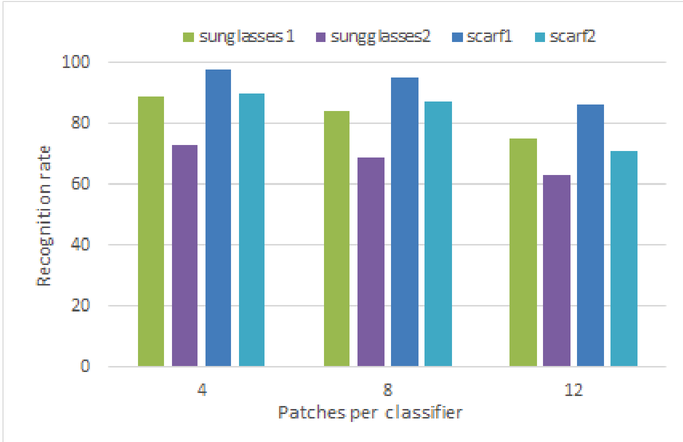

Fig. 9. Recognition performance using different numbers of patches per classifier

For the sake of comparison, the same experiments were conducted with and without random patching and using a single SVM classifier. The results shown in **table I** clearly prove that our random SVM sub-classifiers' best performance is much higher reaching an outstanding 98% when the lower part of the face is covered. This demonstrates the usefulness of our proposed approach.

TABLE I
COMPARING OUR RANDOM PATCHING APPROACH RESULTS TO RESULTS OBTAINED USING THE SAME STEPS WITHOUT THE PATCHING

Test Conditions	Sunglasses		Scarf	
	session1	session2	session1	session2
Random Patching	89	73	98	92
No patching	95	61	89	50

table II compares our approach against some existing approaches in literature. From the table one can see that our method clearly compares favourably against some of the best performing algorithms. For example, our method achieves 98% performance which matches both [13] and [22] in the scarf occluded set. It is is to be noted that the authors [22] have use more than one training sample unlike our method which uses SSPP.

TABLE II
COMPARING OUR RANDOM PATCHING APPROACH RESULTS TO THE LITERATURE

Test Conditions	Sunglasses	Scarf
Our approach	$73 \sim 89$	$92 \sim 98$
LMA/LMA-UDM [13]	$96 \sim 98$	$97 \sim 98$
DICW [22]	99.5	98

VII. CONCLUSION

In this paper a novel face recognition method is presented. First, the images are divided into a number of non-overlapping patches. Then, LBP is used as a local descriptor followed by a dimensionality using KPCA method. Then, a random patch sampling technique is used to build multiple sub-SVM classifiers. Finally, the results are combined using a union rule. Experiments conducted have shown that our approach outperforms classic global SVM face classifiers when the lower part of the face is missing. In addition, the proposed approach compares favourably against with other similar state-of-the-art methods thus demonstrating its potential especially that it operates in an under-sampled environment. Our future work includes the application of our method on both gait and face combined using more challenging datasets taken in the wild.

ACKNOWLEDGMENT

This publication was made possible by NPRP grant # NPR 8 -140 -2-065 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] T. F. Karim, M. S. H. Lipu, M. L. Rahman, and F. Sultana, "Face recognition using pca-based method," in *2010 IEEE International Conference on Advanced Management Science(ICAMS 2010)*, vol. 3, July 2010, pp. 158–162.

[2] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, Nov 2002.

[3] J. Yang, H. Yu, and W. Kunz, "An efficient lda algorithm for face recognition," in *Proceedings of the International Conference on Automation, Robotics, and Computer Vision (ICARCV 2000)*, 2000, pp. 34–47.

[4] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 117–126, 2003.

[5] K. I. Kim, K. Jung, and H. J. Kim, "Face recognition using kernel principal component analysis," *IEEE signal processing letters*, vol. 9, no. 2, pp. 40–42, 2002.

[6] A. Hadid, J. Ylioinas, and M. B. Lpez, "Face and texture analysis using local descriptors: A comparative analysis," in *2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2014, pp. 1–4.

[7] R. Gottumukkal and V. K. Asari, "An improved face recognition technique based on modular pca approach," *Pattern Recognition Letters*, vol. 25, no. 4, pp. 429–436, 2004.

[8] H. K. Ekenel and R. Stiefelhagen, "Local appearance based face recognition using discrete cosine transform," in *13th European Signal Processing Conference (EUSIPCO 2005), Antalya, Turkey*, 2005.

[9] K. Tan and S. Chen, "Adaptively weighted sub-pattern pca for face recognition," *Neurocomputing*, vol. 64, pp. 505–511, 2005.

[10] Y. Zhu, J. Liu, and S. Chen, "Semi-random subspace method for face recognition," *Image and Vision Computing*, vol. 27, no. 9, pp. 1358–1370, 2009.

[11] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," *Image Processing, IEEE Transactions on*, vol. 18, no. 8, pp. 1885–1896, 2009.

[12] A. Sapkota and T. Boult, "Context-patch for difficult face recognition," in *2012 5th IAPR International Conference on Biometrics (ICB)*, 2012, pp. 59–66.

[13] N. McLaughlin, J. Ming, and D. Crookes, "Largest matching areas for illumination and occlusion robust face recognition," 2016.

[14] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European conference on computer vision*. Springer, 2004, pp. 469–481.

[15] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[16] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[17] Y. Wang and Y. Zhang, "The facial expression recognition based on kpca," in *Intelligent Control and Information Processing (ICICIP), **2010** International Conference on*, 2010, pp. 365–368.

[18] X. Xie, W. S. Zheng, J. h. Lai, and C. Y. Suen, "Restoration of a frontal illuminated face image based on kpca," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 2150–2153.

[19] H. M. Ebied, "Feature extraction using pca and kernel-pca for face recognition," in *2012 8th International Conference on Informatics and Systems (INFOS)*, May 2012, pp. MM–72–MM–77.

[20] G. Guo, S. Z. Li, and K. Chan, "Face recognition by support vector machines," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 196–201.

[21] A. M. Martínez and A. C. Kak, "Pca versus lda," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 2, pp. 228–233, 2001.

[22] X. Wei, C.-T. Li, and Y. Hu, "Face recognition with occlusion using dynamic image-to-class warping (dicw)," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–6.