Facial Range Image Matching Using the Complex Wavelet Structural Similarity Metric

Shalini Gupta, Mehul P. Sampat,
Mia K. Markey, Alan C. Bovik
The University of Texas at Austin,
Austin, TX 78712, USA
shalinig@ece.utexas.edu, mehul.sampat@ieee.org,
mia.markey@mail.utexas.edu, bovik@ece.utexas.edu

Zhou Wang
The University of Texas at Arlington
Arlington, TX 76019, USA
zhouwang@uta.edu

Abstract

We propose a novel 3D face recognition algorithm based on facial range image matching using the complex wavelet structural similarity metric (CW-SSIM) metric. Compared with many existing 3D surface matching methods, CW-SSIM is computationally efficient and is robust to small geometrical distortions. Using a data set that contains 360 3D face models of 12 subjects, we tested the performance of the proposed method and compared it with existing 3D surface matching based face recognition algorithms. Verification and identification performance of each algorithm was evaluated by means of the receiver operating characteristic curve and the cumulative match characteristic curve. Among the algorithms tested, the proposed algorithm based on the CW-SSIM resulted in the best overall performance with an equal error rate of 9.13% and a rank 1 recognition rate of 98.6%, significantly better than all the other algorithms. Besides the introduction of a novel approach for 3D face recognition, this is also the first attempt to expand the application scope of complex wavelet domain similarity measure to range image matching in general.

1. Introduction

Numerous present-day applications including access control, surveillance, criminal justice, and human computer interaction require automatic human identification. Face recognition is one of the most widely investigated biometric techniques for human identification. Face recognition systems are advantageous in that they require less user co-operation than some of the other biometric systems (*e.g.*, fingerprint and iris recognition).

Although human beings have evolved to be significantly adept at recognizing faces, it is a difficult task to automate. Despite considerable progress in two dimensional (2D) face recognition systems based on intensity images, they are inadequate for robust face recognition. Their performance is reported to decrease significantly with varying facial pose and illumination [13]. Three dimensional (3D) face recognition systems are less sensitive to changes in ambient illumination conditions than 2D systems [5]. Three dimensional face models can also be rigidly transformed to a canonical pose and location. Hence, considerable research attention is now being directed toward 3D face recognition.

In the computer vision literature, a 3D object that cannot be recognized as either planar or naturally quadric is referred to as a 'free form' object. The surface of the human face can be regarded as an example of a free form object. Techniques based on free form surface matching are often employed for 3D face recognition. In such techniques, the surface of a human face is represented by means of a collection of points in 3D space and referred to as a 'point cloud'. Correspondence between points on the two surfaces is established and the surfaces are alinged. Similarity between the two surfaces is judged by means of a distance metric.

Alternatively, range images of 3D face models in a canonical frontal pose can be generated, and similarity between the two images is established by means of an image similarity metric. A range image, also referred to as a 2.5D surface or depth map, consists of (x, y) points on a regular rectangular grid. Each (x, y) point is associated with a z value or depth of the point on the surface closest to an acquisition device.

For 3D face recognition systems based on surface matching, the similarity/dissimilarity score between an incoming 'probe' face and each face in a 'gallery,' is employed to index the gallery face closest in appearance to the probe face. Hence, the performance of 3D face recognition systems that employ 3D facial surface matching critically depends on the accuracy and robustness of the metric employed for surface matching.

In this study we propose a novel method for 3D face recognition based on facial surface matching using the recently developed image similarity metric called the complex wavelet structural similarity metric [20]. CW-SSIM was origionally proposed to measure the visual similarity between natural images, but, to the best of our knowledge, has never been used for comparing range images. We demonstrate that the performance of the proposed method is superior in terms of recognition accuracy and computational efficiency than some existing techniques for 3D facial surface matching. These include approaches that employ the mean squared error (MSE) between corresponding depth (z) values, the closest point MSE, and the Hausdorff distance.

2. Surface similarity metrics

In this section we review previous work related to 3D face recognition based on surface matching. The MSE metric and the Hausdorff distance have been employed extensively in 3D face recognition algorithms based on surface matching. We provide definitions of these metrics along with that of the CW-SSIM metric.

2.1. Mean squared error

A simple method for comparing two registered range images is by means of the sum of the squares of the Euclidean distances between depth (z) values at corresponding image locations. This simple metric can be referred to as the 'depth MSE'. For two range images A and B, both of size $N \times M$ pixels, the depth MSE (MSE_z) can be calculated as

$$MSE_z = \frac{1}{NM} \sum_{i=1}^{M} \sum_{j=1}^{N} (z_A(i,j) - z_B(i,j))^2. \quad (1)$$

This metric has been employed in previous studies to quantify the distance between coarsely registered range images of 3D human faces [3, 6, 7]. Although this metric is simple to compute and provides tractable solutions when employed in optimization problems, it has been shown to perform poorly for perceptual image quality assessment and pattern recognition problems [19, 16].

Another version of the MSE metric called the 'closest point MSE' has also been employed for 3D facial surface matching. In order to calculate this metric, each 3D surface is regarded as a point cloud. To compute the closest point MSE (MSE_{CP}) between two point clouds $A = \{a_1, a_2, a_3 \dots a_M\}$ and $B = \{b_1, b_2, b_3 \dots b_N\}$, containing M and N points, respectively, we first compute the directed MSE_{CP} from set A to set B as

$$MSE_{CP}(A,B) = \frac{1}{M} \sum_{i=1}^{M} D(a_i, b_{i'}), \quad (2)$$

where D is the square of the Euclidean distance between point $a_i \in A$ and $b_{i'} \in B$ that is closest to the point a_i. The directed $MSE_{CP}(B,A)$ is similarly calculated and the undirected MSE_{CP} between the two surfaces is defined as

$$MSE_{CP} = \max(MSE_{CP}(B,A), MSE_{CP}(B,A)). \quad (3)$$

In numerous previous 3D face recognition studies based on surface matching, the directed MSE_{CP} was employed as the objective function to be minimized in an iterative procedure to align two surfaces as closely as possible [10, 5, 9, 12]. The residual directed MSE_{CP} distance between the two surfaces at the completion of the iterative procedure is used as a measure of similarity between the two registered surfaces for recognition.

2.2. Hausdorff distance

The Hausdorff distance is a measure of similarity between two point sets [4]. The procedure to calculate the Hausdorff distance is similar to that of the MSE_{CP} metric. In order to compute the Hausdorff distance (H) between two point clouds $A = \{a_1, a_2, a_3 \dots a_M\}$ and $B = \{b_1, b_2, b_3 \dots b_N\}$, containing M and N points, respectively, we first calculate the directed Hausdorff distance $h(A,B)$ from set A to set B as

$$h(A,B) = \max_{a \in A} D(a_i, b_{i'}) \quad (4)$$

where D is the square of the Euclidean distance between point $a_i \in A$ and the point $b_{i'} \in B$ that is closest to the point a_i. The directed Hausdorff distance from set B to set A ($h(B,A)$) is similarly calculated and the undirected Hausdorff distance between the two point sets is defined as

$$H = \max(h(A,B), h(B,A)). \quad (5)$$

A variant of the Hausdorff distance is the partial Hausdorff distance. For calculating the partial directed Hausdorff distance from point set A to B, the distances of all points $a_i \in A$ to their closest points in B are sorted in ascending order, and the P^{th} distance in the ordered set quantifies the directed partial Hausdorff distance, *i.e.*,

$$h_P(A,B) = P^{th}_{a \in A} D(a_i, b_{i'}). \quad (6)$$

We can similarly calculate $h_Q(B,A)$, and the undirected partial Hausdorff distance between the two point sets is then defined as

$$H = \max(h_P(A,B), h_Q(B,A)). \quad (7)$$

The partial Hausdorff distance has the advantage of being robust to outliers produced by noise and occlusions. However, its performance depends on the selection of optimal values for the heuristic parameters P and Q that quantify the extent of overlap between the two point sets. Both the Hausdorff distance and the partial Hausdorff distance have been employed previously for measuring the similarity between coarsely alinged 3D face models [1] and as objective functions in an iterative procedure to rigidly align two facial surfaces as closely as possible [11, 15]. Lee and Shim employed a modified Hausdorff distance proposed by Dubuisson and Jain [2], which is closely related to the MSE_{CP} for facial surface matching [8].

2.3. CW-SSIM

Recently, Wang *et al.* proposed the structural similarity metric (SSIM) for predicting human preferences in evaluating image quality [18, 19]. The utility of the metric is not limited to image quality assessment as it can also be employed for pattern recognition tasks. SSIM takes into consideration the local structure and variation about a pixel. It compares structural information, independent of the mean intensity and contrast of the images. It operates in the spatial domain and has been shown to provide good predictions of perceptual image quality for a variety of image distortions [19].

SSIM was extended to the complex wavelet domain resulting in the complex wavelet structural similarity metric [20]. CW-SSIM was demonstrated to be more robust than MSE and SSIM at recognizing hand written digits [20].

CW-SSIM uses the phase information of coefficients in the complex wavelet domain. It is based on the idea that the structural information of image features is mostly contained in the relative phase patterns of wavelet coefficients [20]. To compute the CW-SSIM metric for two images, their complex wavelet transforms are first computed. Let $\mathbf{c}_A = \{c_{A,i} | i = 1, ..., N\}$ and $\mathbf{c}_B = \{c_{B,i} | i = 1, ..., N\}$ be the two sets of coefficients extracted at the same spatial location in the same wavelet subbands of the two images being compared, respectively. The CW-SSIM metric is defined as

$$\tilde{S}(\mathbf{c}_A, \mathbf{c}_B) = \frac{2 \, | \sum_{i=1}^{N} c_{A,i} \, c_{B,i}^{*} | + K}{\sum_{i=1}^{N} |c_{A,i}|^2 + \sum_{i=1}^{N} |c_{B,i}|^2 + K} \, , \qquad (8)$$

where c^* denotes the complex conjugate of c and K is a small positive constant. The CW-SSIM index ranges from a value of 0 to 1, where 1 denotes perfect similarity between two images. A CW-SSIM similarity score s can be converted to a distance measure $d = \sqrt{2(1-s)}$.

For computing the MSE_{CP} and the Hausdorff distance, correspondence between pairs of points on the two surfaces is built by performing an exhaustive search procedure for closest points. This has a high computational complexity of $O((NM)^2)$ for a pair of range images each of size $N \times M$ pixels. This cost can become prohibitive for real-time 3D face surface matching using high resolution data. The computational complexity of the CW-SSIM metric on the other hand is much lower ($O(MN \log_2 MN)$) as it involves calculating complex wavelet coefficients of range images using the FFT algorithm.

The MSE_{CP} and Hausdorff distance metrics are sensitive to alignment between the models being compared. For example, if the models are misaligned, for two points far apart on one surface, their corresponding closest points on the other surface may be very close to each other or may even be the same point. This can lead to erroneous results. In contrast, CW-SSIM is robust to small geometric distortions of the image including small changes in scale, small translations and small rotations [20].

In this study we investigate a novel 3D face recognition algorithm that employs the CW-SSIM metric to compare range images of 3D face models that are coarsely registered. We compare the performance of this metric to the MSE_z, MSE_{CP} and partial Hausdorff distance metrics.

3. Materials and methods

Three dimensional face models for the study were acquired by an MU-2 stereo imaging system by 3Q Technologies Ltd. (Atlanta, GA). The system simultaneously acquires both shape and texture information. The data contained 360 models of 12 subjects. It was partitioned into a gallery set containing one image each of the 12 subjects with a neutral expression. The probe set contained 348 images of the gallery subjects with a neutral or an arbitrary expression. The probe set contained 29 range images of each subject.

Face models were rigidly transformed to frontal orientation using an iterative procedure. Range images were constructed by orthographic projection of the 3D models onto a regularly spaced rectangular grid. The tip of the nose of each model was placed approximately at the center of the image. It should be noted that face surfaces in this study were only coarsely registered by virtue of being in a canonical frontal pose and location. Additional steps to finely align pairs of face surfaces were not performed.

The range images were of size 751×501 pixels with a resolution of 0.32 mm in the x, y, and z directions. Range images were median filtered with a square window of size 3×3 pixels to remove spike noise, interpolated via a process of cubic interpolation to remove large holes and smoothed by applying a Gaussian window with $\sigma = 1$ of size 7×7 pixels. Figure 1 presents example range images employed for the study after the preprocessing steps had

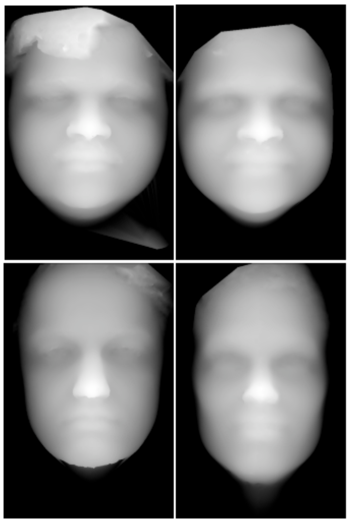

Figure 1. The figure shows examples of the range images that were employed for the study. The images have been preprocessed to remove noise and holes. The two images in the top row are of the same subject.

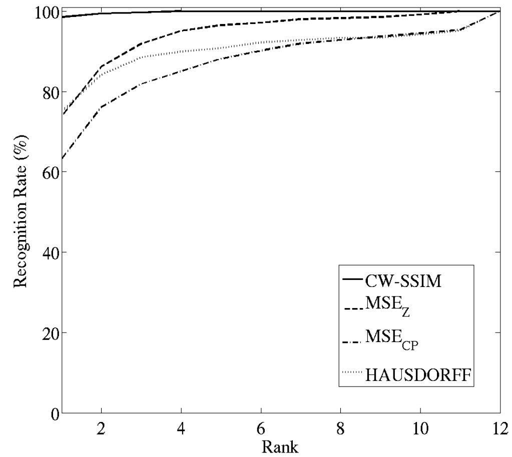

Figure 2. CMC curves for the identification performance of 3D face recognition algorithms based on surface matching that employ different metrics.

been applied.

We implemented four 3D face recognition algorithms based on facial surface matching. The first one employed the CW-SSIM metric to obtain similarity scores between pairs of range images that were converted into distance values. To implement the CW-SSIM index, we first decomposed the range images using a complex version [14] of a 6-scale, 16-orientation steerable pyramid wavelet transform [17]. This is a type of redundant wavelet transform that avoids aliasing in the subbands. The CW-SSIM indices were then computed locally in the complex wavelet domain using a sliding window of size 7×7 pixels.

For the second algorithm, distances between pairs of range images were obtained using the MSE_z metric. For the third and the fourth algorithms, 3D faces were regarded as point clouds and distances between pairs of point clouds were quantified by (a) the MSE_{CP} distance, and (b) the partial Hausdorff distance with $P = Q = 0.9$. In order to reduce the computation time of calculating the MSE_{CP} and Hausdorff distances to a tractable amount, we reduced the size of the range images to 10% of their original size.

Verification performance of the four face recognition algorithms was evaluated using the receiver operating characteristic (ROC) methodology. The equal error rate (EER) and the area under the ROC curve (AUC) for each algorithm was noted. Identification performance was evaluated by means of a cumulative match characteristic (CMC) curve and the rank 1 recognition rate (RR) was observed. The 95% confidence intervals for each observed quantity was obtained by applying a procedure of bootstrap sampling to the outputs of each algorithm. Performance of each algorithm was evaluated separately for the entire probe set, for neutral probes only and for expressive probes only.

4. Results

Equal error rates, AUC values and rank 1 recognition rates of the four 3D face recognition algorithms are presented in Table 1. ROC curves of the four 3D face recognition algorithms are presented in Figure 3. The CMC curves are presented in Figure 2.

The proposed algorithm that employed CW-SSIM for matching 3D facial range images, performed considerably better at identifying human subjects, than the algorithms based on the other metrics. It had a rank 1 $RR = 98.6\%$, $CI = [97.1 \quad 99.7]$ for all probe images. Among the other 3D facial surface matching techniques that were implemented, the overall identification performance of the algorithm that used MSE_z was slightly better than the one that used the partial Hausdorff distance metric (Figure 2). The algorithm that employed the MSE_{CP} metric for surface matching performed the worst (rank 1 $RR = 63.2\%$ for all probes).

Analogous to the identification performance, the verification performance of the algorithm based on the CW-SSIM metric was superior to all the other algorithms with $EER = 9.13\%, CI = [7.71 \quad 10.5]$ for all probes (Fig-

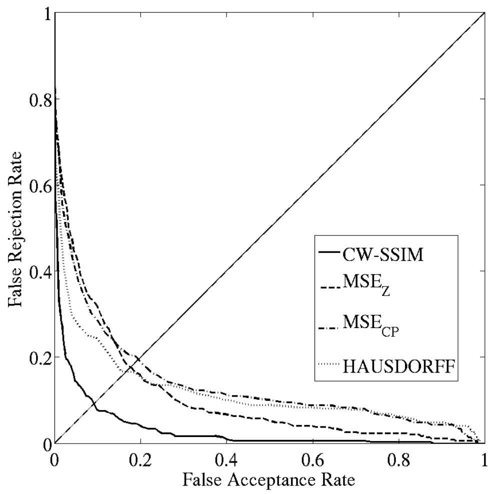

Figure 3. ROC curves depicting the verification performance of 3D face recognition algorithms based on surface matching that employ different metrics.

ure 3). AUC values of the CW-SSIM algorithm for both neutral and expressive probes was significantly lower than the AUC values for the other algorithms (Table 1). The performance of algorithms that employed the MSE_z or the partial Hausdorff distance were not statistically significantly different, with $EER = 17.7\%$ and $EER = 16\%$, respectively, for all probes. Their AUC values for all probes were also not statistically significantly different (Table 1). The highest EER of 19.5% for all probe images was observed for the algorithm that employed the MSE_{CP} metric.

Overall, the proposed algorithm based on the CW-SSIM metric for matching 3D facial range images performed significantly better than the existing 3D facial surface matching algorithms based on the MSE_z, MSE_{CP} and Hausdorff distance.

5. Discussion and conclusions

In this paper we proposed a novel 3D face recognition algorithm based on range image matching using the CW-SSIM metric. We demonstrated that the proposed algorithm is more accurate and robust than some existing face recognition algorithms. The range images that we employed in this study contained 3D models that were only coarsely registered by placing 3D models in a canonical position. The success of the CW-SSIM metric for matching such images can be attributed in part to the fact that CW-SSIM is robust to small geometric distortions including small translations and rotations [20]. Furthermore, the metric is tailored to capture the local structure about a pixel irrespective of the local contrast or luminance values.

Not only is the CW-SSIM metric more accurate at recognizing faces, it also introduces substantial savings in computation time for a 3D face recognition system. Firstly, computing CW-SSIM between a pair of range images/3D models is computationally much less expensive than computing either the MSE_{CP} or the Hausdorff distance between a pair 3D models. Secondly, since CW-SSIM is robust to small image translations and rotations, it does not require that every time a probe is presented to the gallery, it be finely registered to every model in the gallery before the metric can be reliably computed as is the case with some existing 3D facial surface matching techniques [9, 15].

Although in this study the similarity between 3D facial surfaces was quantified using the CW-SSIM metric, the results are not limited only these surfaces. The study points towards the potential applicability of the CW-SSIM metric to other 3D pattern matching tasks as well. One limitation of the technique however, is that in its current form, the metric can only be applied to range images. Hence for matching 3D objects, range images of coarsely registered objects would first have to be created.

In conclusion, we proposed a novel algorithm for 3D face recognition based on range image matching using the CW-SSIM metric. The algorithm was observed to be more accurate, robust and efficient than some existing 3D face recognition algorithms that employ 3D surface matching.

Acknowledgments

We would like to gratefully acknowledge Advanced Digital Imaging Research, LLC, League City, TX, for providing financial support and 3D facial data for the study.

References

[1] B. Achermann and H. Bunke. Classifying range images of human faces with hausdorff distance. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 809–813 vol.2, 2000.

[2] M.-P. Dubuisson and A. Jain. A modified hausdorff distance for object matching. In *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 566–568 vol.1, 1994.

[3] G. G. Gordon. Face recognition based on depth maps and surface curvature. In *SPIE Geometric methods in Computer Vision*, volume 1570, pages 234–247, 1991.

[4] D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):850–863, 1993.

	Neutral Probes		Expressive Probes		All Probes	
Algorithm	RR (%)	CI	RR (%)	CI	RR (%)	CI
CW-SSIM	98.6	[96.8 100]	98.5	[96.2 100]	98.6	[97.1 99.7]
MSE_z	71.3	[65.3 77.3]	78.0	[70.4 84.8]	73.8	[69.1 79.0]
MSE_{CP}	63.0	[56.5 69.0]	63.6	[54.5 72.0]	63.2	[58.0 68.2]
HAUSDORFF	75.0	[69.7 80.1]	75.0	[67.4 81.8]	75.0	[70.1 79.3]
	EER (%)	CI	EER (%)	CI	EER (%)	CI
CW-SSIM	8.29	[6.19 10.3]	9.37	[7.13 12.5]	9.13	[7.71 10.5]
MSE_z	18.7	[16.5 22.2]	15.5	[12.5 18.9]	17.7	[14.7 19.4]
MSE_{CP}	20.6	[16.6 23.0]	18.2	[14.9 22.1]	19.5	[16.7 22.0]
HAUSDORFF	17.0	[14.1 21.8]	15.7	[12.4 19.4]	16.0	[15.5 19.3]
	AUC $\times 10^{-2}$	CI	AUC $\times 10^{-2}$	CI	AUC $\times 10^{-2}$	CI
CW-SSIM	2.87	[1.88 3.74]	3.29	[2.06 5.60]	3.11	[2.21 3.99]
MSE_z	11.2	[9.28 13.4]	10.1	[6.75 14.0]	10.7	[8.56 12.6]
MSE_{CP}	14.1	[10.8 18.2]	12.7	[8.77 17.2]	13.7	[11.4 16.7]
HAUSDORFF	12.1	[9.30 17.1]	11.3	[8.18 16.2]	11.9	[10.1 14.7]

Table 1. The observed rank 1 RR, EER, and AUC values and their 0.025 and 0.975 quantiles for the verification and identification performance of the various 3D face recognition algorithms that were implemented.

[5] E. P. Kukula, S. J. Elliott, R. Waupotitsch, and B. Pesenti. Effects of illumination changes on the performance of geometrix facevision/spl reg/ 3d frs. In *Security Technology, 2004. 38th Annual 2004 International Carnahan Conference on*, pages 331–337, 2004.

[6] S. Lao, Y. Sumi, M. Kawade, and F. Tomita. 3d template matching for pose invariant face recognition using 3d facial model built with isoluminance line based stereo vision. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 911–916 vol.2, 2000.

[7] Y. Lee and T. Yi. 3d face recognition using multiple features for local depth information. In M. Grgic and S. Grgic, editors, *Proceedings EC-VIP-MC 2003. 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications (IEEE Cat. No.03EX667)*, volume vol.1, pages 429–434. Faculty of Electrical Eng. & Comput, Zagreb, 2003.

[8] Y.-H. Lee and J.-C. Shim. Curvature based human face recognition using depth weighted hausdorff distance. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 3, pages 1429–1432 Vol. 3, 2004.

[9] X. Lu, A. K. Jain, and D. Colbry. Matching 2.5d face scans to 3d models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(1):31–43, 2006.

[10] G. Medioni and R. Waupotitsch. Face modeling and recognition in 3-d. In *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, pages 232–233, 2003.

[11] G. Pan, Z. Wu, and Y. Pan. Automatic 3d face verification from range data. In *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*, volume 3, pages III–133–6 vol.3, 2003.

[12] T. Papatheodorou and D. Rueckert. Evaluation of automatic 4d face recognition using surface and texture registration. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 321–326, 2004.

[13] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone. Frvt 2002: Overview and summary. available at www.frvt.org, March 2003.

[14] J. Portilla and E. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–71, 2000.

[15] T. D. Russ, M. W. Koch, and C. Q. Little. A 2d range hausdorff approach for 3d face recognition. In *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, volume 3, pages 169–169, 2005.

[16] P. Simard, Y. Le Cun, J. Denker, and B. Victorri. Transformation invariance in pattern recognition: tangent distance and propagation. *International Journal of Imaging Systems and Technology*, 11(3):181–197, 2000.

[17] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger. Shiftable multiscale transforms. *Information Theory, IEEE Transactions on*, 38(2):587–607, 1992.

[18] Z. Wang and A. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.

[19] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[20] Z. Wang and E. Simoncelli. Translation insensitive image similarity in complex wavelet domain. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, volume 2, pages 573–576, 2005.