A Manifold Approach to Face Recognition from Low Quality Video Across Illumination and Pose using Implicit Super-Resolution

Ognjen Arandjelović
Trinity College
University of Cambridge
Cambridge, CB2 1TQ
oa214@eng.cam.ac.uk

Roberto Cipolla
Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ
cipolla@eng.cam.ac.uk

Abstract

We consider the problem of matching a face in a low resolution query video sequence against a set of higher quality gallery sequences. This problem is of interest in many applications, such as law enforcement. Our main contribution is an extension of the recently proposed Generic Shape-Illumination Manifold (gSIM) framework. Specifically, (i) we show how super-resolution across pose and scale can be achieved implicitly, by off-line learning of subsampling artefacts; (ii) we use this result to propose an extension to the statistical model of the gSIM by compounding it with a hierarchy of subsampling models at multiple scales; and (iii) we describe an extensive empirical evaluation of the method on over 1300 video sequences – we first measure the degradation in performance of the original gSIM algorithm as query sequence resolution is decreased and then show that the proposed extension produces an error reduction in the mean recognition rate of over 50%.

1. Introduction

Identification of humans from low quality video, such as surveillance footage, is of great interest in law enforcement applications. However, it is also an extremely challenging task for computer-based, automatic methods. The reason is that in this setup the most reliable biometrics, such as iris images or fingerprints, are not possible to obtain. In the vast majority of cases, it is only less discriminative modalities, such as face, gait pattern [6] or body appearance (i.e. clothing) [16] that are readily available. Since robust recognition using any of these in isolation has proven very difficult, a number of algorithms take on a multi-modal approach combining e.g. gait and face [19] or gait and height [4], to increase recognition accuracy.

Of unimodal biometrics, face recognition shows the most promise in this setup. Specifically, a vast amount of research has been done on increasing the robustness of face recognition algorithms to the most significant sources of intra-personal appearance variation, such as illumination and pose. However, the most successful of these approaches cannot be readily applied to low resolution data which clearly motivates the use of super-resolution techniques in the pre-processing stages of recognition.

Problem statement. Typically, it is not difficult to obtain a single higher quality video of a person of interest, for the purpose of database enrollment. On the other hand, it is novel imagery that is often of low quality e.g. extracted from surveillance footage. The problem we consider is thus that of matching a single, low resolution video sequence of a moving face (the *query*) to a single, high resolution sequence corresponding to an enrolled person.

Solution summary. The key limitation of previous work that our method addresses is that of smoothing caused by a model-based prior. Our algorithm uses face-specific constraints to formulate a generative model that allows for separation of illumination-affected and downsampling-affected appearance. The latter is shown to be person-specific, and being dependent only on the identity and pose, it can be learnt using a single motion sequence in arbitrary illumination. We propose a statistical model, which is then simply integrated in the adopted Generic Shape-Illumination framework of Arandjelović and Cipolla [1].

Paper organization. The remainder of this paper is organized as follows. In the next section we cover relevant previous work, focusing on super-resolution techniques and their use in visual recognition problems. The original Generic Shape-Illumination (gSIM) algorithm is briefly reviewed in Section 3. The main contribution of this paper is introduced in the section that follows: we start by describing the main

idea of implicit-super resolution in Section 4, mathematically formalize the model for matching sequences in fixed low-resolution in Section 4.1 and, finally, extend the framework to variable resolution matching in Section 4.2. Empirical evaluation of the proposed method is the topic of Section 5: a description of datasets and methodologies is followed by a report of the result and a discussion in Section 5.1. The paper is concluded in Section 6 with a summary of the main contributions and an outline of promising directions for future research.

2. Previous Work

In this section we focus on previous work on super-resolution and, in particular, super-resolution for face recognition; recent general face recognition reviews can be found in [15, 24]

Broadly speaking, super-resolution concerns the problem of reconstructing high-resolution data from a single or multiple low resolution observations. Formally, the process of making a single observation can be written as the following generative model:

$$\mathbf{x} = \downarrow [t(\hat{\mathbf{x}}) + \mathbf{n}], \qquad (1)$$

where $\hat{\mathbf{x}}$ is the high-resolution image, $t()$ an appearance transformation (e.g. due to pose or illumination change, in the case of face images), $\mathbf{n}$ additive noise and $\downarrow$ the down-sampling operator.

In its simplest form, super-resolution takes on the form of non-uniform interpolation: the low resolution input is assumed to correspond to non-identically sampled signal and the high resolution output is produced by interpolative resampling [8, 13, 14]. This approach requires accurate image registration, which is difficult to achieve in the case of faces in extreme illumination conditions.

In the context of the method proposed in this paper, probabilistic approaches, which super-resolve low-resolution images by imposing a class-specific model-based prior are more attractive. Capel and Zisserman [7], for example, use an eigenpart-based prior, applied to six non-overlapping face segments with a Maximum Likelihood estimator. A similar, but holistic approach was proposed by Gunturk *et al.* [9]. Jia *et. al* [11, 12] employ a tensor space prior, but remove the need for manual registration by coupling face registration and super-resolution. The method of Baker and Kanade [3] uses richer, appearance and derivative-based features, but suffers from similar limitations as the previous methods, just as does [18]. As the authors note, the nature of these approaches limits them to frontal faces and non-variable illumination [3].

A number of optical flow-based algorithms were proposed to deal with the problem of facial non-planarity and pose changes [2, 25]. Being highly sensitive to the accuracy of the optical flow estimate, their performance rapidly worsens as the face size is decreased. Another influential group of super-resolution methods use projection onto convex sets (POCS) [17, 20]. These use a families of convex set constraints, one for each image pixel, and an iterative estimation procedure that accounts for object motion, sensor blur and the effects of downsampling. Finally, a number of frequency domain-based techniques have been proposed [10, 21], but these have limited use in the context of very low resolution face images.

In summary, the problem of super-resolution from low quality video footage of faces under variable and arbitrary illumination, and loosely constrained pose is still a major research challenge.

3. The gSIM Algorithm

The Generic Shape-Illumination algorithm of Arandjelović and Cipolla [1] performs face recognition by extracting and matching sequences of faces from unconstrained head motion videos and is robust to changes in illumination, head pose and user motion pattern.

One of the key novelties in the gSIM algorithm is the form in which a learnt prior is applied. Most methods previously proposed in the literature approach the problem from a generative point of view and attempt to learn the distribution of a specific discriminative parameter, e.g. albedo, across human faces. The prior is applied by producing an illumination-normalized representation of a face, which is then matched against those in the gallery. In contrast, gSIM does not explicitly compute such a representation; rather, the prior takes on the form of an estimate of the distribution of non-discriminative, *generic*, appearance changes caused by varying illumination. This is important as it means that unnecessary smoothing of person-specific, discriminative information is avoided.

Specifically, given two images, $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, of a face in the same pose but illuminated differently, gSIM uses a mixture model $\mathcal{G}(\mathbf{d}; \boldsymbol{\Theta})$ to learn the distribution of $\mathbf{d}$, where:

$$\mathbf{d} \equiv \Delta \log \mathbf{x} = \log \mathbf{x}^{(1)} - \log \mathbf{x}^{(2)}. \qquad (2)$$

Under the very general assumption that the mean energy of light incident on the camera is proportional to the face albedo at the corresponding point, $\mathbf{d}$ is approximately generic i.e. not dependent on the person's identity.

This model is employed in recognition by first pose-aligning face sequences that are compared and then computing the model likelihood under the set of observed appearance differences. Pose-wise alignment of faces is achieved using a "reillumination" algorithm: given two face motion sequences, the algorithm produces a third, synthetic one, that corresponds to the first in illumination and to the second in poses. Thus, it can be directly compared with the latter original sequence, producing the likelihood of the same

identity model $\mathcal{G}(\mathbf{d}; \boldsymbol{\Theta})$. The main ideas of the gSIM algorithm are summarized in Fig. 1, while for more detail (and in particular the specifics of achieving pose invariance) we refer the reader to the original paper [1].

4. Implicit Super-Resolution

The original gSIM algorithm relies on the assumption that the appearance difference between two faces in the same pose can be explained by a combination of two factors only: differing identities and differing illumination conditions. However, this is not the case when dealing with real images, as spatial discretization differently affects the appearance of a face at different scales. Gradual image quality degradation as face size is reduced is qualitatively and quantitatively illustrated on an example in, respectively, Fig. 2 (a) and 2 (b).

Unlike previously proposed methods (see Section 2), we propose not to explicitly compute super-resolution face images from low resolution input; rather, we formulate the image formation model in such a way that the effects of illumination and spatial discretization are approximately mutually separable. Thus, we show how the two can be learnt in two stages: (i) a generic illumination model is estimated from a small training corpus of different individuals in varying illumination, and (ii) a low-resolution artefact model is estimated on a person-specific basis, from an appearance manifold corresponding to a single sequence compounded with synthetically generated samples.

4.1. Fixed resolution artefact model

As in the original gSIM algorithm, we assume that the mean energy of light incident at a particular point in the camera plane is a linear function of the albedo a_j of the corresponding point of the imaged face:

$$\hat{x}_j = a_j \cdot s_j \tag{3}$$

where $\mathbf{s}$ is a function of illumination, shape and other parameters not modelled explicitly [1]. However, it is not $\hat{x}_j$ that we directly observe, but rather a spatially discretized image. We model the process of spatial discretization by compounding the model in (3) with a multiplicative factor r_j:

$$x_j = \hat{x}_j \cdot r_j = a_j \cdot s_j \cdot r_j. \tag{4}$$

Now, consider how the introduced modification affects the expression in (2):

$$\Delta \log x_j = \log s_j^{(2)} + \log r_j^{(2)} - $$
$$\log r_j^{(1)} - \log s_j^{(1)} = \Delta \log s_j + \Delta \log r_j \tag{5}$$

It can be seen that the observed appearance difference between two pose-normalized images of the same face at different scales, can now in principle correspond to an arbitrary change in illumination, due to its dependency on the discretization factors too. Unlike the purely illumination-affected component $\Delta \log \mathbf{s}$, the discretization component $\Delta \log \mathbf{r}$ is not generic and the combined effects cannot be learnt using the method proposed in [1].

Let us now consider a high-resolution image of a face $\mathbf{x}$ and its downsampled version $\hat{\mathbf{x}}$. Using the Fourier decomposition of $\mathbf{x}$ we can write:

$$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{x}^{(h)}, \tag{6}$$

where $\mathbf{x}^{(h)}$ is the high-pass filtered $\mathbf{x}$, using the cutoff frequency equal to half the sampling frequency of $\hat{\mathbf{x}}$. Combining the expression (6) and the generative model in (4) gives:

$$\log r_j = \log \frac{x}{\hat{x}_j} = \log \left(1 - \frac{x_j^{(h)}}{\hat{x}_j} \right). \tag{7}$$

Note that this expression depends only on the ratio $x_j^{(h)}/\hat{x}_j$, which is the so-called *local intensity-scaled high-pass filter*, similar to the quasi illumination invariant self-quotient image [23]. Thus, $\log r_j$ is illumination free and only dependent on the identity (for a particular head pose).

4.1.1 Learning the artefact model

We exploit the result that $\log \mathbf{r}$ is not dependent on the illumination using a statistical approach. Specifically, we assume that for person i, the downsampling factor $\log \mathbf{r}$ is distributed across pose according to an underlying probability density function $p^{(i)}(\log \mathbf{r})$ and that it is drawn independently for each video. This allows us to obtain samples from each $p^{(i)}$ given a single sequence per person enrolled in the gallery:

1. samples$^{(i)} = \emptyset$

2. **for** all frames $\mathbf{f}$

3. $\hat{\mathbf{f}} = \text{downsample}(\mathbf{f})$

4. samples$^{(i)} = $ samples$^{(i)} \cup \left(\log \mathbf{f} - \log \hat{\mathbf{f}} \right)$

5. **end**

In this paper we represent each $p^{(i)}$ as a Gaussian density $p^{(i)} = p(\log \mathbf{r}; \boldsymbol{\Theta}_{Art}^{(i)})$, where $\boldsymbol{\Theta}_{Art}^{(i)} = [\mathbf{m}, \mathbf{C}]^T$ are the corresponding model parameters.

(a) Learning the shape-illumination model

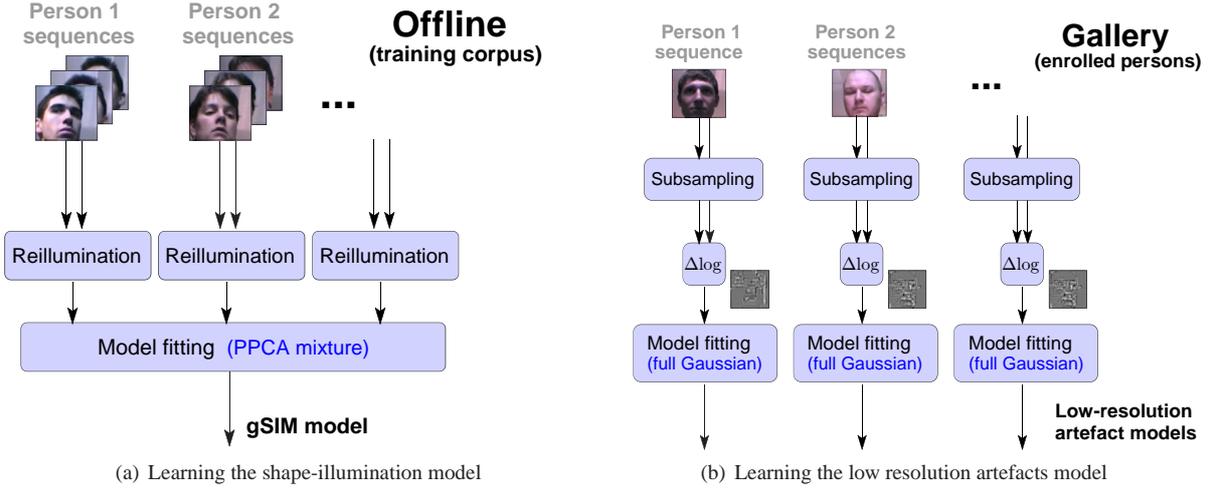(b) Learning the low resolution artefacts model

Figure 1. *A summary of the proposed algorithms for parameters estimation for the (a)illumination and (b) subsampling artefact models. The former is generic and the learning is thus performed offline, using an offline training corpus; the latter are illumination-free and person-specific, and are learnt using the enrolled persons' video sequences.*
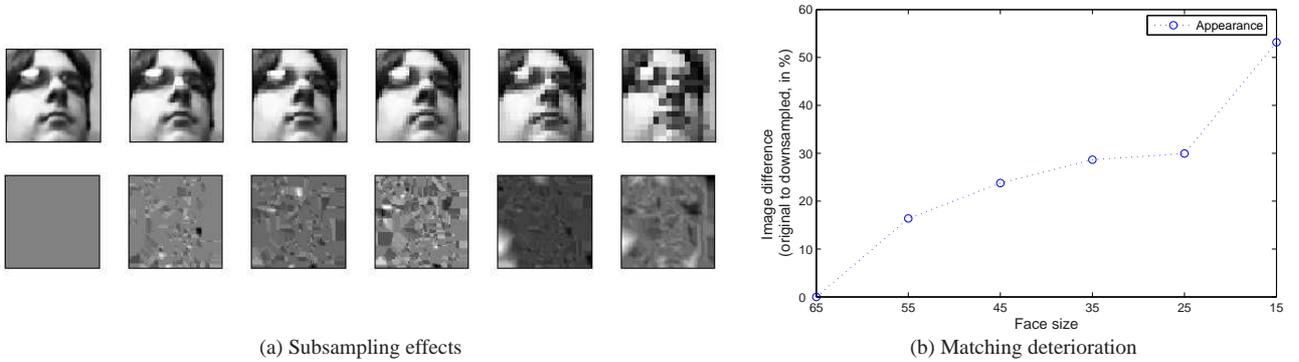

(a) Subsampling effects

(b) Matching deterioration

Figure 2. *Under the piece-wise constant image model, gradual downsampling of a face produces differences which are smooth and progressively greater in magnitude, but of a constrained nature. One of the main results of this paper is that the constraints can be formulated to be illumination-free and person-specific and can thus be learnt using a single exemplar video sequence of an enrolled person.*

4.1.2 Applying the artefact model

We have previously shown in (5) that the log difference of a high and a low resolution appearance image for the same person can be explained by an arbitrary illumination change, given an appropriate downsampling factor. Under the assumption of a uniform prior on the identity over gallery individuals, the likelihood of the same identity of the query and a particular gallery set is proportional to:

$$p(\Delta \log \mathbf{x}) = \int p(\Delta \log \mathbf{x} - \mathbf{r}; \mathbf{\Theta}_{gSIM}) \, p(\mathbf{r}; \mathbf{\Theta}_{Art}) \, d\mathbf{r}, \tag{8}$$

which is the convolution of the two densities

$$p(\Delta \log \mathbf{x}) = p(\Delta \log \mathbf{x}; \mathbf{\Theta}_{gSIM}) * p(\Delta \log \mathbf{x}; \mathbf{\Theta}_{Art}), \tag{9}$$

with the generic illumination effects represented by an N-component mixture of Probabilistic PCA [22] as in [1]. As the convolution of two normal densities is another normal density, $p(\Delta \log \mathbf{x})$ is also an N-component Gaussian mixture, with the k-th component parameters:

$$\hat{\alpha}_k = \alpha_k \tag{10}$$

$$\hat{\mathbf{m}}_k = \mathbf{m}_k + \mathbf{m} \tag{11}$$

$$\hat{\mathbf{C}}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^T + \rho \mathbf{V}_k \mathbf{V}_k^T + \mathbf{C} \tag{12}$$

where $\mathbf{U}_k$ and $\mathbf{\Lambda}_k$ are the principal subspace and the corresponding covariance matrix, $\mathbf{V}_k$ the complementary sub-
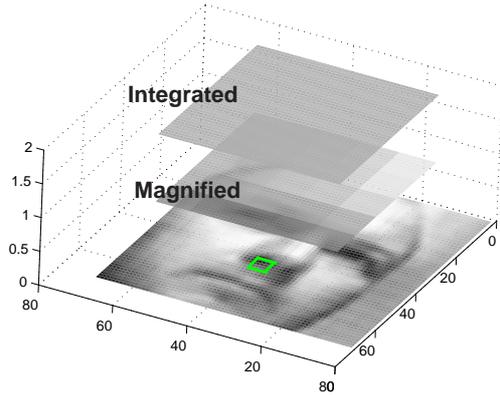
Figure 3. *An illustration of our physical model of downsampling effects. The greyscale pixel value of a low resolution image is proportional to the integral of the high resolution appearance over a square neighbourhood.*

Figure 4. *A schematic summary of the proposed framework for matching of arbitrary resolution input.*

space, and α_k and $\mathbf{m}_k$ the prior and the mean of the k-th Generic Shape-Illumination mixture component.

It follows that the matching stage of our algorithm is exactly the same as in the original gSIM algorithm.

4.2. Variable resolution matching

In the previous section we showed how to implicitly implement super-resolution by learning the person-specific subsampling artefacts at a fixed scale, low-resolution input. We now propose an extension to this framework, that allows for arbitrary low-resolution queries to be matched with high-resolution data.

Consider the physical process of image formation using a CCD: the greyscale intensity value of a particular pixel is proportional on the mean energy of incident light. This can be seen as integration (or averaging) of a square neighbourhood centred at the pixel, as in Fig. 4. As the sampling frequency is reduced, the neighbourhood boundaries are changed *smoothly*. Coupled with the empirical observation that face appearance is also mostly smooth [5], this implies that the effects of downsampling on a piece-wise constant observed appearance are smooth too. This motivates the use of an interpolative extension to our fixed low resolution matching.

Specifically, we learn a hierarchy of downsampling models, which are chosen to be logarithmically equidistant in scale so as to obtain a higher interpolation accuracy at very low resolutions. A novel query sequence is then used to compute the same-identity likelihoods using the artefact models closest to it but lower in scale and closest to it but higher in scale. The likelihood at the scale of query data is estimated by a weighted combination of the two, as illustrated in Fig. 4.
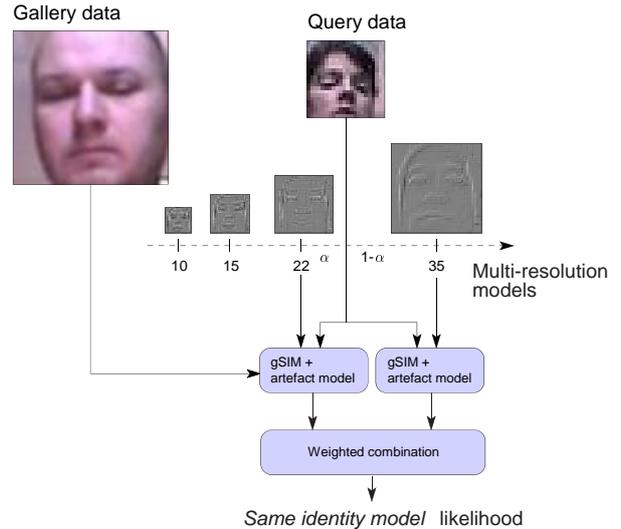
5. Empirical Evaluation

To systematically test the main premises and theoretically derived hypotheses postulated in the preceding sections, we performed empirical evaluation in three stages:

1. In the first set of experiments, we evaluated the rate of performance degradation with decreased query data resolution of the original gSIM algorithm;

2. In the second set of experiments, we evaluated the fixed resolution artefact model;

3. In the final set of experiments, we evaluated our solution to the arbitrary resolution input.

Data sets. We conducted the experiments on two large data sets, the Cambridge Face Database (*CamFace*) and the Toshiba Face Database (*ToshFace*)[1]. These data sets contain 160 individuals and 1300 video sequences, in total, each sampled at 10fps and 10s in length, with extreme illumination and pose changes, and unconstrained head and body motion patterns – the reader is referred to the original publication for a detailed description [1].

Methodology. In all experiments, recognition was performed by matching a single query video sequence to each of the enrolled (or gallery) individuals. Low resolution query sequences were obtained from high resolution data by downsampling. Only a single gallery training sequence was used per enrolled person to learn the person-specific artefact models. To ensure that enough training data was

available for model parameter estimation in the presence of face detector noise and a variable number of detections, we also synthetically enriched the training corpus by adding in (i) mirrored faces and (ii) faces produced by small random translations of the original detections.

5.1. Results and Discussion

A summary of the results of the three experiments is shown in Tab. 1. As predicted by our image formation model (4), the first set of experiments demonstrated a significant performance degradation of the original gSIM algorithm when scale normalization is performed using simple pixel-space interpolation as proposed in [1]. At the 50×50 pixel scale the recognition rate is nearly perfect across the two data sets. Downsampling to 40% of the original scale, or 20×20 pixels, increased the average error rate to 8.9%. This is already a sufficiently high error rate that it is unlikely that a multi-modal approach employing other weakly discriminative biometrics would greatly increase its usability as a security system. With further downsampling the error rate was rapidly increased even more, averaging 18.2% for 15×15 pixel faces and 28.0% for 10×10 pixels. These results are interesting in the light of the performance of the gSIM algorithm reported in the original publication. Although the authors used simple image resampling to normalize for scale differences, the recognition was virtually unaffected due to little scale variation in the data, as illustrated in Fig. 5. In the context of our artefact model, this means that $p^{(i)}(\log \mathbf{r})$ were close to delta functions and the convolution in (9) simply resulted in an unchanged generic effects model. Finally, it is interesting to note a far greater increase in the error rate of the gSIM algorithm on the Cambridge database, compared to the Toshiba dataset. At this stage we can propose no plausible mechanism that would explain this observation.

In the second set of experiments, we evaluated the effectiveness of the proposed implicit super-resolution algorithm using a single set of fixed resolution artefact models, first at 20×20 pixel scale and then at 10×10 pixels. On both databases, the recognition performance was vastly improved, more than halving the previously measured error rates. A similar result was obtained in the last set of experiments in which we tested the proposed method for interpolating model likelihoods at the scale of 15×15 pixels. While this demonstrates its effectiveness compared to the original method, further experiments are needed to fully understand the relationship between artefact model distances in scale, the quality of likelihood interpolation and the recognition performance of our algorithm.
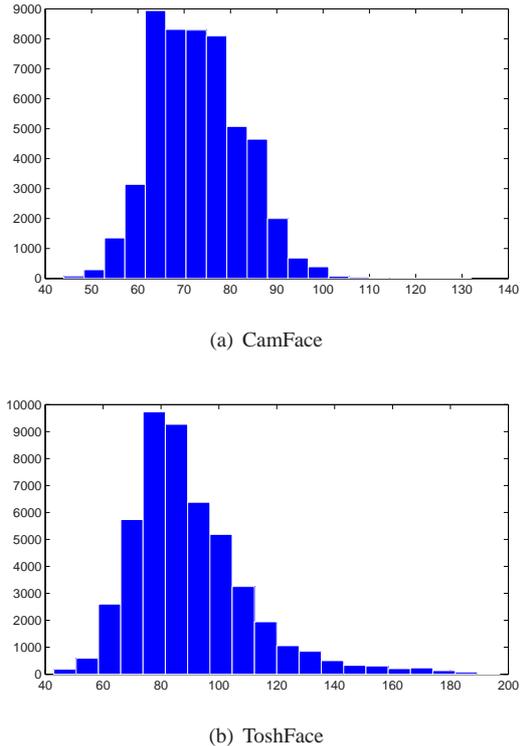

(a) CamFace

(b) ToshFace

Figure 5. *Histograms of sizes of detected faces in pixels (square bounding box) for CamFace and ToshFace datasets.*

6. Summary and Conclusions

In this paper we introduced a method for face recognition that matches low resolution query video against gallery sequences which are of much higher quality. The main novelty is the formulation of a generative model that allows for separation of illumination and downsampling effects, and the residual, discriminative appearance. Specifically, the illumination effects are shown to be generic across faces and are thus learnt offline from a small training corpus. On the other hand, the effects of downsampling in our model are person-specific and illumination-free and are learnt using enrolled persons' training sequences. The power of this approach lies in the observation that all prior information we use is inherently non-discriminative and recognition is achieved by *implicit* application of super-resolution. Thus, our method of does not exhibit the same level of discriminative power loss as do algorithms which smooth out data by explicitly reconstructing super-resolved imagery. Preliminary empirical results on a large data set confirm these premises.

Acknowledgements

Table 1. *Average error rates (%) and the corresponding standard deviations across data sets.*

Database	No. of sequences	Query face size (pels)	Proposed ISR–gSIM	gSIM, subsampled then scaled
CamFace	700	10	12.2 / 2.1	36.6 / 7.5
		15	6.2 / 1.1	25.3 / 5.5
		20	4.2 / 0.8	12.1 / 4.5
		50	n/a	0.3 / 0.8
ToshFace	600	10	8.6 / 1.7	13.9 / 4.4
		15	2.8 / 1.0	6.5 / 2.4
		20	2.0 / 0.7	3.5 / 1.1
		50	n/a	0.1 / 0.5
Total	1321	10	10.8 / 2.0	28.0 / 6.5
		15	5.0 / 1.1	18.2 / 4.6
		20	3.4 / 0.8	8.9 / 3.6
		50	n/a	0.2 / 0.7

search.

References

[1] O. Arandjelović and R. Cipolla. Face recognition from video using the generic shape-illumination manifold. *In Proc. European Conference on Computer Vision (ECCV)*, 4:27–40, May 2006. 1, 2, 3, 4, 5, 6

[2] S. Baker and T. Kanade. Investigation into optical flow super-resolution for surveillance applications. *Technical Report CMU-RI-TR-99-36, Robotics Institute, Carnegie Mellon University*, 1999. 2

[3] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(9):1167–1183, 2002. 2

[4] C. BenAbdelkader, R. Cutler, and L. Davis. Person identification using automatic height and stride estimation. *In Proc. IEEE International Conference on Pattern Recognition (ICPR)*, 4:377–380, 2002. 1

[5] M. Bichsel and A. P. Pentland. Human face recognition and the face image set's topology. *Computer Vision, Graphics and Image Processing: Image Understanding*, 59(2):254–261, 1994. 5

[6] I. Bouchrika and M. S. Nixon. People detection and recognition using gait for automated visual surveillance. *In Proc. IEE International Symposium on Imaging for Crime Detection and Prevention*, 2006. 1

[7] D. P. Capel and A. Zisserman. Super-resolution from multiple views using learnt image models. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001. 2

[8] J. J. Clark, M. R. Palmer, and P. D. Lawrence. A transformation method for the reconstruction of functions from non-uniformly spaced samples. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(4):1151–1165, 1985. 2

[9] B. K. Gunturk, Y. Batur, A. U. Altunbasak, M. H. Hayes, and R. M. Mersereau. Eigenface-domain super-resolution for face recognition. *IEEE Transactions on Image Processing*, 12(5):597–606, 2003. 2

[10] T. S. Huang and R. Tsai. Multi-frame image restoration and registration. *Advances in Computer Vision and Image Processing*, 1:317–339, 1984. 2

[11] J. K. and S. Gong. Multi-resolution patch tensor for facial expression hallucination. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:395–402, 2006. 2

[12] J. K., S. Gong, and A. Leung. Coupling face registration and super-resolution. *In Proc. IAPR British Machine Vision Conference (BMVC)*, 2006. 2

[13] S. P. Kim, N. K. Bose, and H. M. Valenzuela. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38:1013–1027, 1990. 2

[14] T. Komatsu, K. Aizawa, T. Igarashi, and T. Saito. Signal-processing based method for acquiring very high resolution image with multiple cameras and its theoretical analysis. *In Proc. of Institution of Electrical Engineers*, 140:19–25, 1993. 2

[15] S. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi. Recent advances in visual and infrared face recognition – a review. *Computer Vision and Image Understanding (CVIU)*, 97(1):103–135, 2005. 2

[16] C. Nakajima, M. Pontil, and T. Poggio. People recognition and pose estimation in image sequences. *International Joint Conference on Neural Networks (IJCNN)*, 4:189–194, 2000. 1

[17] M. I. Sezan. An overview of convex projections theory and its application to image recovery problem. *Ultramicroscopy*, 40:55–67, 1992. 2

[18] O. Sezer, Y. Altunbasak, and A. Ercil. Face recognition with independent component based super-resolution. *In Proc. of SPIE Visual Communications and Image Processing Conference*, 2006. 2

[19] G. Shakhnarovich, L. Lee, and T. Darrel. Integrated face and gaitrecognition from multiple views. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:439–446, 2001. 1

[20] H. Stark and Y. Yang. *Vector Space Projections: A numerical Approach to Signal and Image Processing, Neural Nets, and Optics.* Wiley, 1998. 2

[21] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan. High resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. *In Proc. IEEE International Conference on Acousics, Speech and Signal Processing*, 1:317–339, 1992. 2

[22] M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999. 4

[23] H. Wang, S. Z. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. *In Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 819–824, 2004. 3

[24] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2004. 2

[25] W. Zhao and H. S. Sawhney. Is super-resolution with optical flow feasible? *In Proc. European Conference on Computer Vision (ECCV)*, pages 599–613, 2006. 2