Gradient Edge Map Features for Frontal Face Recognition under Extreme Illumination Changes

Ognjen Arandjelović Swansea University, UK
ognjen.arandjelovic@gmail.com

Abstract

Our aim in this paper is to robustly match frontal faces in the presence of extreme illumination changes, using only a single training image per person and a single probe image. In the illumination conditions we consider, which include those with the dominant light source placed behind and to the side of the user, directly above and pointing downwards or indeed below and pointing upwards, this is a most challenging problem. The presence of sharp cast shadows, large poorly illuminated regions of the face, quantum and quantization noise and other nuisance effects, makes it difficult to extract a sufficiently discriminative yet robust representation. We introduce a representation which is based on image gradient directions near robust edges which correspond to characteristic facial features. Robust edges are extracted using a cascade of processing steps, each of which seeks to harness further discriminative information or normalize for a particular source of extra-personal appearance variability. The proposed representation was evaluated on the extremely difficult YaleB data set. Unlike most of the previous work we include all available illuminations, perform training using a single image per person and match these also to a single probe image. In this challenging evaluation setup, the proposed gradient edge map achieved 0.8% error rate, demonstrating a nearly perfect receiver-operator characteristic curve behaviour. This is by far the best performance achieved in this setup reported in the literature, the best performing methods previously proposed attaining error rates of approximately 6–7%.

1 Introduction

The aim of this work is to match images of frontal faces across extreme illumination changes. This is a problem of importance in a broad range of practical applications. Indeed, in security applications which authenticate the user before granting access to a particular resource, the user is asked to face the camera. Examples include passport checks, entry control to buildings and mobile phone authentication, amongst others. Person based retrieval systems working on highly unconstrained data, such as that extracted from TV films and series, also often focus on nearly frontal faces [3, 9] in no small part because face detection is most reliable for this pose. Frontal faces can then be synthesized from non-frontal views. Everningham and Zisserman [9] achieve this using a generic 3D head model, Gross *et al.* [12] adopt an active appearance model while Wong *et al.* [23] describe a regression based method.

 http://dx.doi.org/10.5244/C.26.12

The process of imaging a face is inherently lossy. In projecting the 3D shape of a face onto a 2D surface, discriminative depth information is lost [18]. This problem is particularly pronounced when faces are imaged in the frontal pose. A convincing corpus of evidence both from computer vision as well as human physiology supports the observation that the frontal pose is not optimal for recognition [13, 16]. The key reason for this is that in this case the most salient facial features extend towards the camera, making their depth variation hard to judge. Therefore it is unsurprising that even the best performing methods of today struggle when applied in a setting in which only a single image of a face is used for training and another single image as a query, with possibly extreme illumination changes between the two. The challenge is readily seen in Figure 1 which shows some of the illumination conditions which motivate the present work. Cast shadows, the presence of overhead (elevation/inclination up to 90°), lateral and even rear (azimuth up to 130°) illumination, and overexposed facial regions are just some of the difficulties of note.

Figure 1: Examples of extreme illuminations present in the data set used in this paper.

In this paper we describe an algorithm which comprises a cascade of processing steps, each aimed either at harnessing further discriminative information or at normalizing for a particular source of extra-personal appearance variability. The end result of the cascade is an image which can be easily matched against other images computed in the same way. This representation retains discriminative information contained around characteristic facial features and, considering that the image plane 2D geometry is preserved, the spatial relationship between these features.

2 Computing Invariant & Discriminative Representation

In this section we introduce the main contribution of the present paper – a cascade of processing steps which produce a discriminative and invariant representation of a person's face. We begin by looking at some of the representations which are widely used in face recognition and motivate the proposed approach by highlighting their limitations in the presence of extreme illumination conditions. This is a followed by detailed explanation of the cascade which is for the convenience of the reader also summarized in the diagram in Figure 2.

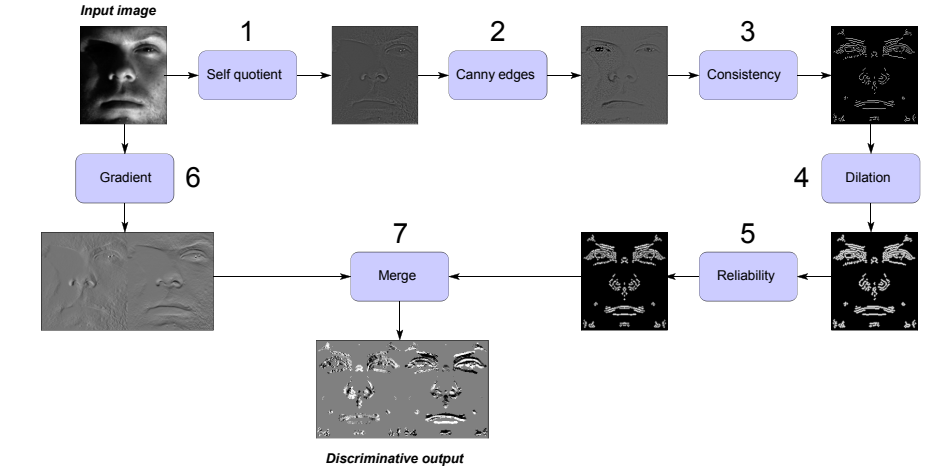


Figure 2: An overview of the key steps of the proposed dual pipeline cascade used to produce an invariant and discriminative representation of a face.

2.1 The Cascade: Step by Step

Person specific, discriminative information is not uniformly distributed across different parts of a face. Rather, the areas around characteristic facial features, such as the eyes and the nose are generally the most important ones for recognition. Unsurprisingly, this observation is exploited by most automatic methods. Algorithms based on elastic graph matching [6, 7, 14] or local feature analysis [9], amongst others, explicitly use the appearance of only a select number of characteristic features. Others adopt one of a variety of statistical approaches [1, 5, 19, 20, 22] and when applied in a discriminative setting end up automatically learning exactly these characteristic features as the ones most useful for recognition.

Most of the discriminative information is contained in the parts of the face exhibiting substantial variation in either geometry (that is to say, facial surface normal direction) or albedo. By the very nature of the imaging process, these variations generally produce regions of observed appearance variation in images too. As such, they can be readily detected using direct computations on pixel intensities. One of the simplest methods of accomplishing this is by applying a 2D high pass filter. Generally, the filter is applied in the spatial domain by first convolving the image I with a Gaussian kernel G to produce the low pass image I_{LP}:

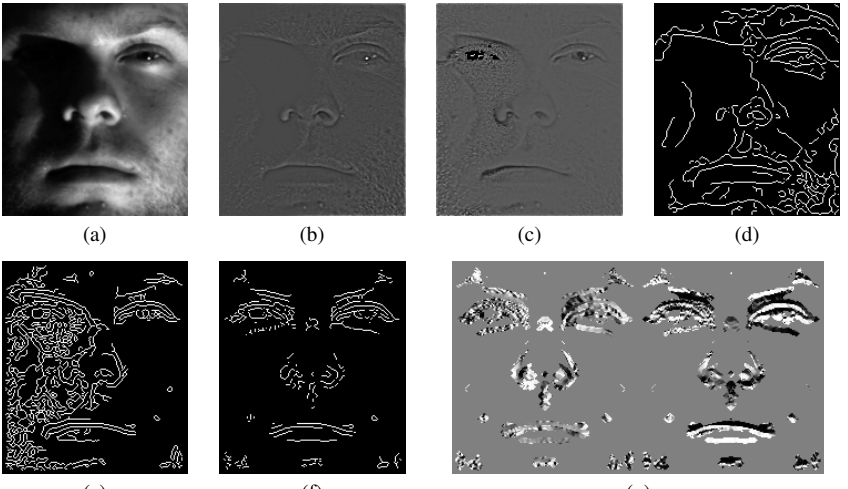
$$I_{LP} = I * G \tag{1}$$

and then subtracting the low pass image from I:

$$I_{HP} = I - I_{LP} \tag{2}$$

This filter has been widely used in face recognition. For example, Fitzgibbon and Zisserman show that it can be used to suppress the effects of illumination for face clustering in films [10]. However, when applied on images of faces under extreme illuminations such as those considered in this paper, the simple high pass filter fails in achieving a satisfactory result. As our experiments in Section 3 will demonstrate, the error rate of 68.49% for raw images is reduced only to 47.98% by high pass filtering. One of the reasons for this poor

performance can be readily observed by examining a representative input image, such as that shown in Figure 3(a) and the high pass filtered result in Figure 3(b). Specifically, notice that the discontinuities in the poorly lit, shadowed regions are much less pronounced than those in well illuminated regions.



(a) (b) (c) (d) (e) (f) (g)

Figure 3: Representations produced at different stages of the proposed pipeline: (a) original raw appearance, (b) band pass filtered appearance, (c) self-quotient image, (d) Canny edges computed from the original image, (e) Canny edges computed from the self-quotient image, (f) symmetrically consistent and reliable edges, and (g) the final representation of gradient directions constrained to the neighbourhood of symmetrically consistent and reliable edges.

The dependence of the magnitude of the intensity discontinuities preserved by high pass filtering can be addressed by using one of the Retinex-like methods first described by Land [15]. These methods are in part inspired by the human visual system and the observation that humans perceive brightness in a relative rather than absolute manner. In other words, a discontinuity of a small magnitude in a dark region should have a greater effect than a discontinuity of the same magnitude in a bright region. The high pass filter can this be modified simply by dividing pixel-wise the filtered result with the low pass filtered image which has the effect of averaging image intensity. This is a variant of the self-quotient image [21]:

$$I_{SQI}(x,y) = \frac{I_{HP}(x,y)}{I_{LP}(x,y)} \tag{3}$$

The result of applying this filter on the same input image as before is shown in Figure 3(c) which indeed appears to be an improvement over the output of the high pass filter. However, when this representation is used for matching on our data set, the outcome is perhaps surprising. As we will discuss in Section 3, the error rate on our data set is in fact increased to nearly that achieved by using raw appearance. A more detailed inspection of the image in Figure 3(c) reveals insight into the causes. Specifically, the noise in the poorly lit regions of the face has been amplified, as has the originally imperceptible boundary of the shadow caused by (weak) ambient illumination. The filter also causes the appearance of artefacts around interfaces between very bright and very dark image regions.

Steps 1 and 2: Provisional Edges Our method avoids the described difficulties associated with the use of absolute intensity by concentrating on the binary edge image. This is the first step of the proposed cascade. Note that we do not detect edges directly in the original image. Instead, we apply the Canny edge detector on the self-quotient image, to ensure that very weak edges in poorly illuminated regions are correctly detected. The difference between the two approaches is illustrated in Figure 3(d) and Figure 3(e). Note that our approach results in higher Canny thresholds (automatically estimated based on the histogram of the input image gradients). While this has the effect of producing fewer false edges in the well lit regions of the face and more true edges in the poorly lit regions, the number of spurious edges in the poorly lit regions is also increased. This problem is addressed in the next step of our cascade.

Step 3: Spurious Edge Removal The edge map computed from the self-quotient image may contain many false edges. Some of these may result from the amplification of noise in poorly lit regions of the face. Others may be strong edges at the boundaries of cast shadows. Highly saturated image regions may also cause the hallucination of edges. Regardless of what the underlying cause is, false edges can decrease the matching accuracy. For example, it is straightforward to see that the left hand side of the edge map in Figure 3(e), full of densely packed false edges, will match nearly any true face edge map rather well.

We remove most false edges by exploiting the vertical symmetry of frontal faces. Specifically, we require a degree of agreement between the left hand and right hand sides of the edge map. If E is the binary edge image and $\overset{\frown}{\cdot}$ the vertical mirroring operator the new binary edge image E_T with spurious edges removed is computed as follows:

$$E_T(x,y) = \begin{cases} 1 & : E(x,y) = 1 \text{ and } \overset{\frown}{E}_{DT}(x,y) \leq 2 \\ 0 & : \text{otherwise} \end{cases} \tag{4}$$

where $E_{DT} = f_{DT}(E)$ is the distance transformed edge image. In other words, we remove all edge segments which are not within ≈ 2 pixels from the corresponding mirrored edges. An example result is shown in Figure 3(f).

Steps 4 and 5: Edge Reliability Refinement Note that after spurious edges are removed in the previous step of the cascade, the resulting binary image E_T is not necessarily vertically symmetric. Since a perfectly invariant representation of a frontal face should be vertically symmetric, we interpret this lack of symmetry as arising from true but unreliable edges, i.e. edges which are not repeatably detected across different illuminations.

To ensure that the final representation contains only those true edges which are repeatably detectable, we again exploit the vertical symmetry of frontal faces. We first dilate the edge image E_T using a 4×4 pixel sized solid circle structuring element S and then combine the dilated edge information from the left hand and right hand sides of the face:

$$E_R(x,y) = \min \left[E_T(x,y) \oplus S, \overset{\frown}{E}_T(x,y) \oplus S \right] \tag{5}$$

where $\oplus$ is the dilation operator.

Steps 6 and 7: Merging Edge and Gradient Information In the last step of the proposed cascade, we incorporate into our representation further discriminative information. The specific limitation of the edge map that we wish to overcome is its limited ability to robustly

capture shape. This is a consequence of the observation that each edge map pixel by itself only contains information about whether an edge passes through it or not. Edge pixels carry no additional information about the directionality of the corresponding edge. We demonstrate that a highly discriminative representation can be obtained by combining the dilated reliable edges map and the corresponding gradient phase. This is achieved by computing a 3D image comprising two "stacked" 2D images which contain horizontal and vertical gradient information at the image loci at which the dilated edge map is non-zero:

$$E_{GM}(x,y,1) = \begin{cases} \frac{\partial I}{\partial x}/|\nabla I| & : E_R(x,y) > 0 \\ 0 & : E_R(x,y) = 0 \end{cases} \tag{6}$$

$$E_{GM}(x,y,2) = \begin{cases} \frac{\partial I}{\partial y}/|\nabla I| & : E_R(x,y) > 0 \\ 0 & : E_R(x,y) = 0 \end{cases} \tag{7}$$

Notice that the horizontal and vertical gradients are normalized by the magnitude of the total gradient. This is performed with the intention of taking into account the unreliability of absolute or even relative image intensity across different illuminations. The directionality of the gradient, on the other hand, is preserved well in the vicinity of strong discontinuities (but not necessarily elsewhere). The proposed representation is illustrated in Figure 3(g), displayed as the two stacked images side by side.

3 Evaluation

We evaluated the proposed representation on the YaleB database [11]. This is a challenging data set used as a standard benchmark for the comparison of face recognition algorithms in terms of their robustness to severe illumination changes. The variation in the direction of the dominant light source illuminating a face is extreme: its azimuth varies from -130° to 130°, and its elevation from -40° (i.e. pointing upwards) to 90° (i.e. directly overhead, pointing downwards), giving a total of 64 different illumination conditions. Notice that the face is sometimes illuminated from the rear lateral direction (and thus hardly illuminated at all), that extreme cast shadows are often present as are highly bright saturated image regions. Some of these challenges have already been illustrated in Figure 1. The database does not include any intentional variation in facial expression, but some variation exists nonetheless, mainly in the form of squinting when the subject is facing the dominant light source.

It is important to emphasize just how much more challenging our evaluation protocol is in comparison to those adopted by most of the previous work on this database. The first major difference is the range of illumination variation, which is severally constrained in most previous work. For example, Georghiades *et al.* [11] evaluate their method on the subset of the database which includes only those images for which the angle between the dominant light direction and the viewing direction is at most 45°. This has the effect of halving the number of images used and removes exactly the most challenging illumination conditions. Another significant difference is that we use only a single image for training and only a single probe image for testing the algorithm. In contrast, the method of Georghiades *et al.* requires seven images with specific dominant illumination directions for training. Lee *et al.* [17] use nine training images. Yet other methods require a manual placement of predetermined characteristic feature points, such as the method of Xie and Lam [24] which uses twenty such points. Evaluated in the same setup as ours, the highest reported rates achieved on this database do not exceed approximately 93–94% [8].

Table 1: Mean recognition error rates achieved using different representations: raw appearance, high pass filtered appearance (HPF), self quotient image (SQI), an edge image, a Gaussian blurred edge image, and the proposed gradient edge map. In all cases recognition was assessed in a 1–to–N matching scenario, using a single training image per person and a single probe image.

	Raw	HPF	SQI	Edges	Blurred edges	Proposed
All conditions	0.6849	0.4798	0.6297	0.2801	0.1757	0.0082
"Easy" subset	0.4881	0.2513	0.1432	0.1260	0.0203	0.0000

Unlike most of the previous work, we evaluate the effectiveness of the proposed representation using all the available illuminations, including the most extreme ones. To obtain an estimate of the mean 1–to–N recognition rate, we perform 10,000 simulated recognition challenges, randomly choosing a single training image per person. A test, probe image is then matched against all of them and assigned to the nearest class using the nearest angle distance. In other words, if $f_R(\ldots)$ is a function which rasterizes an image, the similarity between a probe image I' and a training image I is computed as:

$$\rho(I',I) = \frac{f_R(E'_{GM})^T f_R(E_{GM})}{|f_R(E'_{GM})||f_R(E_{GM})|} \tag{8}$$

where E'_{GM} and E_{GM} are the gradient edge map representations computed from I' and I. Using the same matching method, we also obtain the receiver-operator characteristic curves.

3.1 Results

The key recognition results are summarized in the first row of Table 1 which shows the average recognition error achieved using different representations of a face. As expected, unprocessed appearance performs poorly, making an incorrect decision in 68.49% of the cases. High pass filtering improves the results somewhat, reducing the error rate by 70% down to 47.98%. This trend is consistent with the previous reports in the literature.

On this data set the self quotient image representation does not fare as well as the high pass filter. As explained in Section 2, this is a consequence of several factors. Firstly, notice the amplification of quantum noise in the shadowed regions. This is noise which is caused by the randomness in the exact number of photons hitting a particular photosensitive element and is especially pronounced if the average number of photons is low i.e. if the corresponding region is dark. In addition, self quotient image filtering increases quantization noise. In dark image regions, relative brightness is not captured with good accuracy, as quantization is coarser. Lastly, the severe illumination conditions used in the acquisition of many of our images tend to result in the creation of artefacts in self quotient images. Specifically this occurs at the boundaries of often sharp contrast between shadowed and well lit regions of the face that are created in severe illuminations.

To substantiate our theoretical explanation of the relatively poor performance of the self quotient image, we also performed an additional set of experiments, using only an "easy" subset of data which constrains both the azimuth and the elevation of the dominant light direction to $\pm 20^\circ$ deviation from the camera direction. The results are summarized in the second row of Table 1. Unsurprisingly, higher recognition rates are achieved for all representations. What is interesting to observe is that in this experiment self quotient image performs

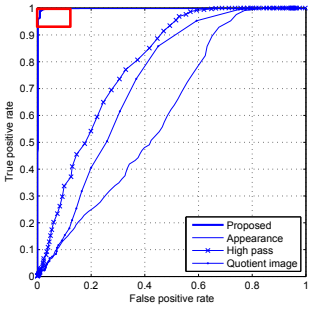
much better than the high pass filtered appearance. While the error rate was previously increased by 31% with contextual contrast normalization, it is now decreased by 43%, which is a remarkable difference in behaviour. These results serve to illustrate why neither of the two representations, the high pass filtered appearance or the self quotient image, are appropriate for recognition in extreme illumination conditions, motivating the ideas introduced in the present paper (also see [2]).

We also investigated the performance attained using a simple binary edge map, considering its role in the derivation of the proposed representation. Evaluated both on the entire data set and on the "easy" subset, the edge map outperformed both the high pass filter and the quotient image. The effect was more pronounced in extreme illumination conditions. The performance was improved further yet when we applied a Gaussian blur on the edge map, thereby increasing the robustness of the representation to precise localization of edges and registration errors. The additional improvement was particularly noticeable on the "easy" data subset. This suggests that an edge map extracts rather discriminative, meaningful information and that the majority of matching errors when this representation is employed under mild illumination conditions comes precisely as a consequence of small misalignments. On the other hand, the relatively poor absolute performance of the edge map in challenging conditions reinforces our argument that the presence of cast shadows, saturated image regions and noise is so substantial that a more complex and robust representation is needed for accurate recognition.

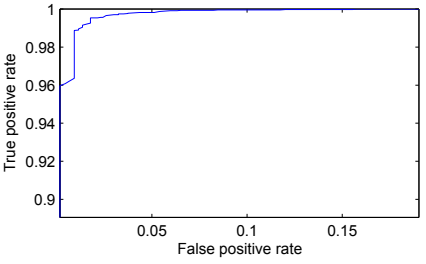
Finally, we evaluated our gradient edge map representation. Both on the "easy" subset and on the entire data corpus, the proposed representation achieved outstanding results, matching flawlessly in the former case and with 0.8% error rate in the latter. This is by far the best result reported in the literature to date. The best performing methods described previously which use only a single training image and a single probe image, evaluated on the entire range of illuminations present in YaleB database achieve rates of 93–94% [8]. This means that the error rate attained with the proposed representation is reduced approximately eightfold. This exceptional performance of the gradient edge map representation is further corroborated by the receiver-operator characteristic curves, computed across the entire YaleB data set, shown in Figure 4(a). The characteristic curves of raw appearance, high pass filter and self quotient image based approaches is rather poor, their relative ordering matching that of their mean recognition rates discussed previously. On the same plot, the characteristic curve of our representation is so close to the ideal curve, that it is difficult to distinguish with a naked eye. Therefore in Figure 4(b) we also include a magnification of the salient region near the point corresponding to the zero false positive and 100% true positive rate. From this plot it can be seen that the corresponding equal error rate is approximately 0.8%. In a high security setting, at 0.15% false positive rate, the true positive rate is at a very high 96%.

4 Conclusions

In this paper we introduced a novel gradient edge map representation for frontal face recognition. Our approach was motivated by the limitations of relative image intensity based representations widely used in the literature, but which we show are ineffective in extreme illumination conditions. Instead, the proposed representation relies on the phase of gradient information near repeatably detectable image intensity edges which characterize salient facial features. Repeatably detectable image intensity edges are extracted using a cascade of processing steps, each of which seeks to harness further discriminative information or



(a)



(b)

Figure 4: (a) Receiver-operator characteristic curves achieved using different representations. (b) Magnified receiver-operator characteristic attained using the proposed representation near the point corresponding to the zero false positive and 100% true positive rate. The proposed gradient edge map achieves nearly perfect performance, with the equal error rate of 0.8%.

normalize for a particular source of extra-personal appearance variability. The effectiveness of the proposed representation was demonstrated on the notoriously challenging YaleB data set, which covers a wide range of illumination conditions, many of which are extreme (rear lateral, overhead). Unlike most of the previous work we used only a single image per person for training and a single probe image as test, and did not eliminate any of the images from the evaluation. Our gradient edge map achieved outstanding results, incorrectly recognizing in only 0.8% of the cases and exhibiting nearly perfect receiver-operator characteristic behaviour. This performance vastly exceeds that reported previously in the literature on this data set and using the same evaluation methodology.

Our immediate future work will concentrate on extending the proposed method to deal with varying pose. For example, this may be achievable by applying the gradient edge map features on synthetically generated images of frontal faces.

References

[1] O. Arandjelović. Computationally efficient application of the generic shape-illumination invariant to face recognition from video. *Pattern Recognition (PR)*, 45 (1):92–103, 2012.

[2] O. Arandjelović and R. Cipolla. A new look at filtering techniques for illumination invariance in automatic face recognition. *In Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 449–454, 2006.

[3] O. Arandjelović and R. Cipolla. Automatic cast listing in feature-length films with anisotropic manifold space. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2:1513–1520, 2006.

[4] S. Arca, P. Campadelli, and R Lanzarotti. A face recognition system based on local feature analysis. *Lecture Notes in Computer Science (LNCS)*, 2688:182–189, 2003.

[5] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks (TNN)*, 13(6):1450–1464, 2002.

[6] D. S. Bolme. Elastic bunch graph matching. Master's thesis, Colorado State University, 2003.

[7] B. Duc, S. Fischer, and J. Bigün. Face authentication with Gabor information on deformable graphs. *IEEE Transactions on Image Processing (TIP)*, 8(4):504–516, 1999.

[8] P. Dunker and M. Keller. Illumination normalization for face recognition – a comparative study of conventional vs. perception-inspired algorithms. *BIOSIGNALS*, 2: 237–243, 2008.

[9] M. Everingham and A. Zisserman. Identifying individuals in video by combining generative and discriminative head models. *In Proc. IEEE International Conference on Computer Vision (ICCV)*, 2005.

[10] A. Fitzgibbon and A. Zisserman. On affine invariant clustering and automatic cast listing in movies. *In Proc. European Conference on Computer Vision (ECCV)*, pages 304–320, 2002.

[11] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 23(6):643–660, 2001.

[12] R. Gross, I. Matthews, and S. Baker. Active appearance models with occlusion. *Image and Vision Computing (special issue on Face Processing in Video)*, 1(6):593–604, 2006.

[13] J. H. Hsiao and T. T. Liu. The optimal viewing position in face recognition. *Journal of Vision*, 12(2), 2012.

[14] C. Kotropoulos, A. Tefas, and I. Pitas. Frontal face authentication using morphological elastic graph matching. *IEEE Transactions on Image Processing (TIP)*, 9(4):555–560, 2000.

[15] E. H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–128, 1977.

[16] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju. Finding optimal views for 3D face shape modeling. *In Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 31–36, 2004.

[17] K. Lee, J. Ho, and D. Kriegman. Nine points of light: Acquiring subspaces for face recognition under variable lighting. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:519–526, 2003.

[18] G. Pan and Z. Wu. 3D face recognition from range data. *International Journal of Image and Graphics*, 5(3):573–593, 2005.

[19] O. G. 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.

[20] A. Stergiou, A. Pnevmatikakis, and L. Polymenakos. EBGM vs. subspace projection for face recognition. *In Proc. International Conference on Computer Vision Theory and Applications*, 2006.

[21] 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 (FG)*, pages 819–824, 2004.

[22] Y. Wang, Y. Jia, C. Hu, and M. Turk. Non-negative matrix factorization framework for face recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 19(4):495–511, 2005.

[23] Y. Wong, C. Sanderson, and B. C. Lovell. Regression based non-frontal face synthesis for improved identity verification. *In Proc. International Conference on Computer Analysis of Images and Patterns (CAIP)*, pages 116–124, 2009.

[24] X. Xie and K.-M. Lam. Face recognition under varying illumination based on a 2D face shape model. *Pattern Recognition (PR)*, 38(2):221–230, 2005.