Face Recognition Using the Classified Appearance-based Quotient Image

Masashi Nishiyama and Osamu Yamaguchi
Corporate Research & Development, Toshiba Corporation, Japan
{masashi.nishiyama,osamu1.yamaguchi}@toshiba.co.jp

Abstract

We propose a new method for synthesizing an illumination normalized image from a face image including diffuse reflection, specular reflection, attached shadow and cast shadow. The method is derived from the Self-Quotient Image (SQI) [14] which is defined by the ratio of albedo at the pixel value to a locally smoothed pixel value. However, the SQI is not synthesized from an image containing shadows or specular reflections. Since these regions correspond to areas of high or low albedo, they cannot be discriminated from diffuse reflection by using only a single image. To classify the appearances, we utilize a simple model defined by a number of basis images which represent diffuse reflection on a generic face. Through experimental results we show the effectiveness of this method for face identification on the Yale Face Database B and on a real-world database, using only a single image for each individual in training.

1. Introduction

In face recognition, identification performance is significantly influenced by a variation in appearance caused by illumination. The appearance is mainly classified into four components [11]: diffuse reflection, specular reflection, attached shadow and cast shadow. To increase the accuracy of face recognition in the case of variation of these components, it is important to be able to synthesize an illumination normalized image for training from any given face image. A method for synthesizing such images without an explicit appearance model for illumination has been proposed in [1, 9], where the variation caused by illumination is regarded as an intraclass variation. An illumination normalized image is synthesized by projection onto a feature space which is insensitive to the intraclass variation, whilst remaining sensitive to the interclass variation which represents changes in individual appearance. However, in the case of different lighting conditions amongst images for learning the feature space, an illumination normalized image cannot be synthesized since the feature space cannot estimate a novel appearance caused by illumination.

In order to synthesize an illumination normalized image with an appearance model for illumination, a method using basis images has been proposed in [12, 10], which is suitable for modeling the variation in face appearance caused by diffuse reflection. This model is advantageous in that a registration system stores only a few images and fitting to the model is simple. Shashua et al. have proposed the Quotient Image (QI) [12], which is the ratio of albedo between a face image and linear combination of basis images for each pixel. The ratio of albedo is illumination invariant. However, the QI assumes that all observed image intensities are generated by diffuse reflection. Okabe et al. synthesized an image consisting of diffuse reflection and attached shadow, removing specular reflection and cast shadow from a face image [10]. In this method, diffuse reflection and attached shadow are estimated by basis images using random sample consensus. Basis images were generated from three face images acquired under fixed pose and a moving point light source [11], or from four face images acquired under moving pose and a fixed point light source [8]. Therefore, these methods [12, 10] require multiple face images for each individual in the registration stage of training. This requirement is an important limitation for practical applications.

In order to synthesize an illumination normalized image from a single face image for each individual, Wang et al. have proposed the Self-Quotient Image (SQI) [14]. Each pixel of the SQI contains the ratio of the albedo at that pixel and a smoothed value of the albedo in a local region. However, in the case that a face image contains specular reflection and shadow, the SQI is not illumination invariant. For this reason, diffuse reflection cannot be extracted from the pixel affected by specular reflection and a shadow region cannot be discriminated from a dark albedo region.

In this paper, through introducing a model of basis images to the SQI for the classification of appearances caused by illumination, we propose a new method for synthesizing an illumination normalized image from a face image which includes diffuse reflection, specular reflection, attached shadow and cast shadow. Each individual in a training set shares the same basis images which represent diffuse

reflection on a generic face. The basis images are generated using face images acquired from other individuals in the training stage. We call the synthesized image the Classified Appearance-based Quotient Image (CAQI). To synthesize the CAQI, we design a filter, which extracts the ratio of albedo in a local region, based on the appearance of each pixel. The appearance is classified into four components using the basis images.

The remainder of this paper is organized as follows. First, we describe the conventional method, SQI, and clarify the weaknesses in section 2. Next, we describe a new method, CAQI, in section 3. Then, we demonstrate the effectiveness of CAQI through experiments in section 4.

2. Obtaining invariance with respect to diffuse reflection

2.1. Self-Quotient Image (SQI)

The SQI [14] has been proposed for synthesizing an illumination normalized image from a single face image. The SQI is defined by a face image $I(x, y)$ and a smoothed image $S(x, y)$ as

$$Q(x,y) = \frac{I(x,y)}{S(x,y)} = \frac{I(x,y)}{F(x,y) * I(x,y)} , \qquad (1)$$

where $F(x, y)$ is a smoothing filter; $*$ is the convolution operation. In the case that a smoothing filter is an isotropic Gaussian filter $G(x, y)$, equation (1) is equivalent to the center/surround retinex transform described in [5]. The SQI is generated using an anisotropic, weighted Gaussian filter $W(x, y)G(x, y)$.

As described in [14], $Q(x, y)$ is illumination invariant if certain assumptions are met. The diffuse reflection is defined by the Lambertian model as

$$i = \max(al\mathbf{n}^T\mathbf{s}, 0) , \qquad (2)$$

where i is pixel value of $I(x, y)$; a and $\mathbf{n}$ are the albedo and the normal of the object surface; l and $\mathbf{s}$ are the strength and direction of the light source. The lengths of $\mathbf{n}$ and $\mathbf{s}$ are normalized. The attached shadow appears in the case where the dot-product between $\mathbf{n}$ and $\mathbf{s}$ is negative. **We make two assumptions in a local region** : that (a) l, $\mathbf{n}$, and $\mathbf{s}$ are uniform and (b) all observed appearance is diffuse reflection. Then the ratio of albedo is extracted as

$$Q(x,y) = \frac{a(x,y)l\mathbf{n}^T\mathbf{s}}{F(x,y) * a(x,y)l\mathbf{n}^T\mathbf{s}} = \frac{a(x,y)}{F(x,y) * a(x,y)} . \qquad (3)$$

Under multiple light sources, the ratio of albedo is also obtained using additivity $\mathbf{s} = \sum_{k=1}^{n} \mathbf{s}_k$.

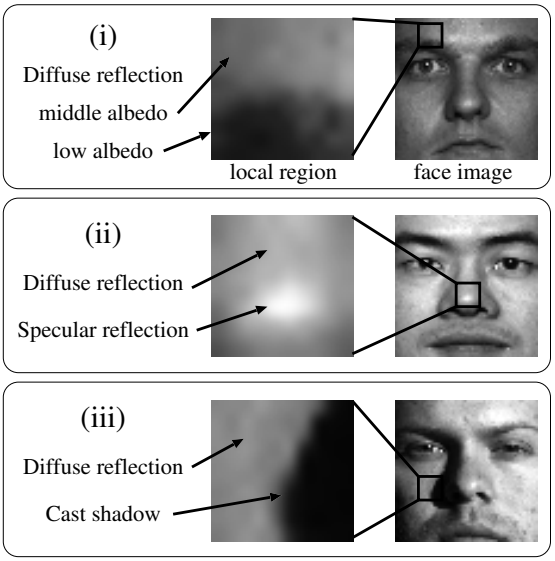

Figure 1. Examples of different appearances in local regions. Other than diffuse reflection is also observed in local regions (ii) and (iii). Then, illumination invariant feature is not extracted from these local regions.

2.2. Problems caused by effects other than diffuse reflection

Equation (3) works well for a local region such as Figure 1(i) which includes low albedo region of eyebrow and middle albedo region of forehead. The assumptions of section 2.1 are valid in this local region where only diffuse reflection is observed. However, the assumptions are violated in (ii) and (iii) where cast shadow, or specular reflection, is partially observed. In specular reflection region, diffuse reflection cannot be extracted because the pixel value is saturated. Then, we need to estimate diffuse reflection value from the pixel giving rise to specular reflection. In cast shadow region, light parameters, l and $\mathbf{s}$, are different from diffuse reflection because cast shadow appears the direct path between a light source and a point obstructed by an intervening object. The same is true of attached shadow region. In these shadow region, diffuse reflection is also observed because the region is illuminated by another light source e.g. ambient light. Then, we need to determine a local region where similar appearance is observed.

2.3. Weight function for synthesizing the SQI

The weight function $W(x, y)$ of the SQI is designed for each pixel to prevent halo effects around edge. The weight

function divides a local region into two subregions M_1 and M_2 . If $I(x, y) \in M_2$ then $W(x, y) = 0$ else $W(x, y) = 1$. The subregion is determined by a threshold τ which is the average of pixel values in the local region. The subregion with the larger number of pixels is set to M_1, the other is set to M_2. However, the weight function is unable to discriminate between the shadow and low albedo region, e.g. eyes and eyebrows. Illumination invariant feature can be extracted from a local region which contains edge between middle albedo region and low albedo region such as Figure 1(i). Therefore, the SQI is disadvantageous in that an important feature representing identity is neglected. Another problem is that a region is not simply divided where pixel values change smoothly, e.g. the soft shadow observed in the contour of the cast shadow.

3. Classified Appearance-based Quotient Image (CAQI)

In this section we introduce a new method for extracting the ratio of albedo from a face image including specular reflection and shadow. In this method the weight function for the weighted Gaussian filter is calculated using classification of appearance for each pixel.

3.1. Classification of appearance caused by illumination using photometric linearization

To classify surface appearance into diffuse reflection, specular reflection, attached shadow and cast shadow, we utilize photometric linearization [7]. Photometric linearization transforms a face image into a linearized image consisting of only diffuse reflection. A linearized image $\tilde{I}(x, y)$ is defined by the model [11] in which arbitrary images caused by diffuse reflection are represented by a linear combination of three basis images $I_i(x, y)(i = 1, 2, 3)$ taken under point light sources in linearly independent directions as

$$\tilde{I}(x, y) = \sum_{i=1}^{3} c_i I_i(x, y) \ . \tag{4}$$

To estimate a coefficient c_i from a face image including specular reflection and shadow, photometric linearization uses random sampling of pixels. By random sampling we calculate candidates of c_i. The final c_i are determined from the distribution of candidates by iterating between outlier elimination and recomputation of the mean.

The appearance is classified using a difference image $I'(x, y)$, defined as

$$I'(x, y) = I(x, y) - \tilde{I}(x, y) \ . \tag{5}$$

As in [7] a pixel having a negative value in $I'(x, y)$ is classified as cast shadow. A pixel having a value greater than the threshold in $I'(x, y)$ is classified as specular reflection. A pixel having a negative value in $\tilde{I}(x, y)$ is classified as attached shadow. To estimate diffuse reflection for the pixel giving rise to specular reflection, we replace the pixel value from $I(x, y)$ to $\tilde{I}(x, y)$.

3.2. Generation of basis images

Since we do not acquire basis images $I_i(x, y)$ for each individual in the training stage, we generate basis images using different individuals from the training stage. For this purpose, we acquire images for multiple persons under fixed pose and a moving point light source not including specular reflection and shadow. The point light source is moved so as to be linearly independent. We apply Singular Value Decomposition (SVD) using all the acquired images. The vectors, selected in descending order of singular value, are the basis images. We assume that the basis images represent the diffuse reflection on a generic face. However, retaining the first three basis images, the estimated c_i in equation (4) has an error induced by the difference between a generic face and an individual face. For fitting basis images to various individuals, we select more than four vectors in descending order of the singular value. These vectors represent the principal component of the variation in appearance due to the diffuse reflection for individuals.

3.3. Calculation of the weighted function using classified appearance

To calculate the weight function based on appearance caused by illumination for the weighted Gaussian filter, we utilize a difference image $I'(x, y)$ representing the difference of the appearance. We aim to extract the ratio of albedo from a local region in which similar appearance is observed. For example, if the center of a local region is classified as diffuse reflection, we give a large weight to diffuse reflection of the surrounding area of the center and small weight to shadow of that. The weight function $W(x, y)$ is defined by comparing the center of a Gaussian filter with its surrounding area as follows:

$$W(x, y) = \frac{1}{1 + \alpha |I'(x, y) - I'(x_0, y_0)|} \ , \tag{6}$$

where (x_0, y_0) is the center pixel of the Gaussian filter; α is constant ($\alpha > 0$). If $I'(x, y)$ is greater than $I'(x_0, y_0)$, then $I'(x, y)$ has a different appearance to $I'(x_0, y_0)$, and a small weight is given to $I'(x, y)$.

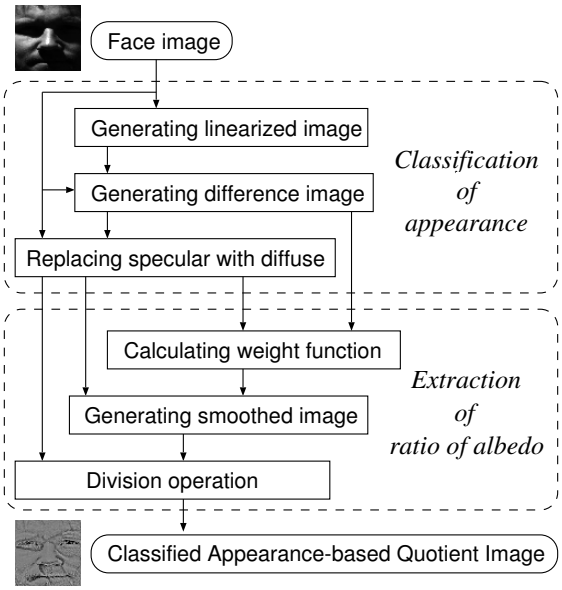

Figure 2. Flow of synthesizing the CAQI.

3.4. Algorithm for synthesizing the CAQI

We now explain the algorithm for synthesizing the Classified Appearance-based Quotient Image (CAQI). The flow diagram of the algorithm is shown in Figure 2. First, a face image is aligned from the positions of the pupils and the nostrils. Next, appearance is classified using the photometric linearization. Then, the specular reflection is replaced with the estimated diffuse reflection.

For calculating the ratio of the albedo, we need to determine the size of a local region in order to obtain a uniform surface normal $\mathbf{n}$. However, searching for a suitable size is difficult since there is an ambiguity in the computation of $\mathbf{n}$ for basis images which are generated by unknown l and $\mathbf{s}$ [3]. Therefore we use multiple sizes as in [14, 6]. The size is defined by the standard deviation σ of the Gaussian filter. We calculate the weight function $W_j(x, y)$ for each $\sigma_j (j = 1, \ldots, N)$. Finally, the illumination normalized image $Q(x, y)$ is synthesized as

$$Q(x,y) = \sum_{j=1}^{N} f\left(\frac{I(x,y)}{W_j(x,y)G_j(x,y) * I(x,y)}\right) . \quad (7)$$

Note that we assume $\int\int W_j(x,y)G_j(x,y)dxdy = 1$. As in [14, 6] we use f in equation (7). The function f prevents asymptotic approach to infinity from division when the term $W_j(x,y)G_j(x,y) * I(x,y)$ tends to zero in shadowed regions.

Figure 3. Examples of the Yale face database B.

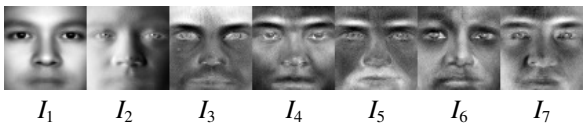

Figure 4. Basis images generated from the CMU-PIE 68 individuals under 8 lighting conditions.

4. Empirical Evaluation

4.1. Performance for varying illumination

Experimented conditions. To illustrate the performance of our proposed method, we have conducted face identification experiments using the Yale Face Database B [2]. The database consists of images taken under 64 different lighting conditions, which are divided into 5 subsets. In Figure 3, we show examples of face images in each subset. We used 640 images in total taken of 10 individuals in a frontal pose. We located a 64×64 pixels face image from the positions of the pupils and the nostrils obtained manually. In our experiments we use a single face image of each individual for training, where the lighting condition is the same for all subjects. This condition is different from [14] which used multiple images for training. The images taken under the remaining 63 lighting conditions are used for testing.

For the generation of basis images, we used the CMU PIE [13] consisting of different individuals from the Yale Face Database B. We selected images taken from a frontal pose (c27) under light sources (f06, f07, f08, f09, f11, f12, f20, f21). We applied SVD to all images of 68 individuals and selected the 7 basis images shown in Figure 4.

Identification performance. We compared the performance of the CAQI with the gray-scale image (GS), the histogram equalized image (HE), the Multi-scale Center/Surround Retinex image (MSR) [6] and the SQI [14]. For GS, we used the face image directly. For HE, we used the histogram equalization of face image. For MSR, SQI and CAQI, we used multiple Gaussian filters, σ_i were set to $1.0, 2.0, 3.0$, and f in equation (7) was chosen as the logarithmic function. For CAQI, we sampled 20,000 candidates

Table 1. CMR (%) on the Yale Face Database B. Images taken under one lighting condition were used for training set. Images taken under 63 different lighting conditions were used for test set.

Subset (Training)	Method				
	GS	HE	MSR	SQI	CAQI
1	68	71	93	87	96
2	64	68	91	82	95
3	54	58	82	71	88
4	39	43	79	68	84
5	25	42	79	75	91

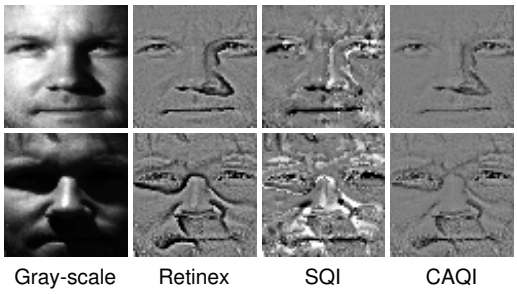

Figure 5. Examples of an image synthesized by each method. (The upper row is subset 3 and the lower row is subset 4)

to estimate c_i and α was set to 0.1. The synthesized image was transformed to a vector by raster-scanning of the image and the length of the vector was normalized. Then, we calculated the similarity between the training vector and the testing vector using normalized correlation.

Table 1 shows the evaluation result for each method in terms of the correct match rate (CMR). CMR is the probability that a face image of the right person is accepted correctly when using nearest neighbor classification. We show the average of the CMR for each subset in the table. The performance of the MSR is superior to that of the SQI. Heusch et al. reported a similar result for a different database [4]. For this reason, we infer that the weight function for the SQI causes the problem that the center pixel of Gaussian filter is not included in the subregion M_1. Since the appearance differs between the center pixel and the subregion, the assumptions for equation (3) are invalid. From the table we can see that the performance of CAQI is superior to that of the others. In particular, the performance in subset 5 where faces are mostly shadowed is improved significantly.

Table 2. Comparison of identification performance using basis images of each individual versus basis images of others. Images taken under one lighting condition in subset 1 were used for training set.

Method	Subset (Input)				
	1	2	3	4	5
SQI	100	97	91	76	70
CAQI-other	100	97	97	93	93
CAQI-same	100	99	99	95	94

Comparison of synthesized images. Figure 5 shows the synthesized center/surround retinex image, SQI and CAQI. To synthesize these images we used a single $\sigma = 1.0$. In the case of the synthesized images in subset 4, features hidden by shadow around the eyes in the gray-scale image appear in the center/surround retinex image, SQI, and CAQI. It can clearly be observed that the influence of cast shadows is reduced in the CAQI. This is particularly evident at the contour of the shadowed region and in the region of the specular reflection on the nose.

Identification performance in the case of basis images for each individual. We compared the identification performance using basis images of each individual (CAQI-same) and using basis images of others (CAQI-other). In CAQI-other, we used basis images of CMU-PIE shown in Figure 4. In CAQI-same, basis images for each individual in a training set were generated using 7 face images in subset 1. We selected three basis images in descending order of singular value. Images taken under one lighting condition in subset 1 were used for training set. Images taken under 63 different lighting conditions were used for the test set. Table 2 shows the evaluation result in terms of CMR. CAQI-other and CAQI-same are superior to the SQI. However, the CAQI-same is superior to CAQI-other. For this reason, we presume that the albedo and the surface normal of faces in the test set are different from those for individuals included in CMU-PIE. To increase performance further, we are considering developing a method of deforming basis images from generic basis images without requiring more than one training image per person.

4.2. Performance assessment on a real-world database

We also evaluated the methods using a database collected under two lighting conditions for 100 individuals. We assumed a practical application of face recognition to passport-based identification. The lighting conditions are (I) no shadow on the face by using a flash attached to cam-

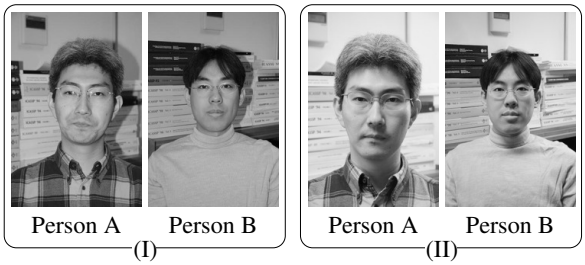

Figure 6. Examples of a real-world database.

Table 3. CMR (%) on a real-world database (100 individuals).

Training image	Method				
	GS	HE	MSR	SQI	CAQI
(I)	30	16	78	65	82
(II)	33	25	73	29	80

era, and (II) face shadowed by a single spotlight mounted on the ceiling as shown in Figure 6. We collected a single face image for each individual and each lighting condition in the case that ceiling fluorescent lamps are on. The parameters of section 4.1 remained unchanged. We show the CMR $(\%)$ in Table 3 for the condition that (I) is the training set and (II) is the testing set, and vice versa.

We can see that the method using the CAQI is superior to the other methods. In particular, the CAQI is effective when images containing cast shadow regions, as in (II), are used as training example. However, an error in the photometric linearization arose since a single set of c_i does not fully represent the diversity of illumination under multiple light sources. To improve the performance, we consider developing a method of estimating multiple sets of c_i under multiple light sources.

5. Conclusion

This paper proposed a method of synthesizing an illumination normalized image, in the case that diffuse reflection, specular reflection, attached shadow and cast shadow are observed, by classifying appearances into four components using basis images which represent the diffuse reflection on a generic face. Our method is able to obtain high identification performance on the Yale Face Database B and on a real-world database, using only a single image for each individual in training.

In future work, we intend to develop a method of estimating basis images for each individual in training from basis images of others, and a method of estimating the sets of coefficients for the combination of basis images under multiple light sources, and a method of determining the size of a local region by adjusting facial normal.

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 19(7):711 – 720, 1997.

[2] 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*, 23(6):643 – 660, 2001.

[3] H. Hayakawa. Photometric stereo under a light source with arbitrary motion. *The Journal of the Optical Society of America A*, 11(11):3079 – 3089, 1994.

[4] G. Heusch, F. Cardinaux, and S. Marcel. Lighting normalization algorithms for face verification. *IDIAP-Com 05-03*, 2005.

[5] D. J. Jobson and Z. Rahman. Properties and performance of a center/surround retinex. *IEEE Transactions on Image Processing*, 6(3):451 – 462, 1997.

[6] D. J. Jobson, Z. Rahman, and G. A. Woodell. A multi-scale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965 – 976, 1997.

[7] Y. Mukaigawa, H. Miyaki, S. Mihashi, and T. Shakunaga. Photometric image-based rendering for image generation in arbitrary illumination. *IEEE International Conference on Computer Vision*, 2:652 – 659, 2001.

[8] A. Nakashima, A. Maki, and K. Fukui. Constructing illumination image basis from object motion. *Proceeding of the 7th European Conference on Computer Vision-Part III*, pages 195 – 209, 2002.

[9] M. Nishiyama, O. Yamaguchi, and K. Fukui. Face recognition with the multiple constrained mutual subspace method. *5th International Conference on Audio- and Video-based Biometric Person Authentication 2005*, pages 71 – 80, 2005.

[10] T. Okabe and Y. Sato. Object recognition based on photometric alignment using ransac. *IEEE Proceeding Conference on Computer Vision and Pattern Recognition*, 1:221 – 228, 2003.

[11] A. Shashua. Geometry and photometry in 3d visual recognition. *Ph. D. Thesis*, 1999.

[12] A. Shashua and T. Riklin-Raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 23(2):129 – 139, 2001.

[13] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615 – 1618, 2003.

[14] H. Wang, S. Z. Li, and Y. Wang. Generalized quotient image. *IEEE Proceeding Conference on Computer Vision and Pattern Recognition*, 2:498 – 505, 2004.