Image Detection Under Varying Illumination and Pose

Margarita Osadchy Daniel Keren
Department of Computer Science
University of Haifa
Haifa 31905, Israel
e-mail: (gamer,dkeren)@cs.haifa.ac.il

Abstract

This paper focuses on the detection of objects with Lambertian surface under both varying illumination and pose We offer to apply a novel detection method that proceeds by modeling the different illuminations from a small number of images in the training set, this automatically voids the illumination effects, allowing fast illumination invariant detection, without having to create a large training set It is demonstrated that the method "fits in" nicely with previous work about the modeling of the set of object appearances under varying illumination In the experiments, an object was correctly detected under image plane rotations in a 45-degrees range, and a wide variety of different illuminations

1. Introduction

Slight changes in pose and illumination produce large changes in object appearance. Recognition of objects under various classes of geometric transformations or under various viewpoints was previously studied in [6, 12,14,16]. However, these methods offer no solution for the problem of illumination variability in natural images. In [1,2,4,13] the problem of varying illumination and fixed pose was addressed. Recognition under large variation in pose and illumination has recently been introduced in [3]. In this method each "cone" [2] models only 4x4 degrees patch of the visibility sphere, hence large variability in pose is accomplished by calculation of the distance to each cone, which is much more computationally expensive than our approach.

Appearance-based methods [6,7,8,10,11,12,14,15,16, 18,19,20] can recognize the object under a particular pose and lighting, if the object has been previously seen under similar circumstances. To extend these methods to handle illumination variability, a large set of images of the object under varying illumination should be used for the learning stage, which is highly inefficient. The following observations [2,5,9,17] allow to alleviate this problem, by modeling the object appearance under a wide range of illuminations, instead of physically creating them.

Consider a convex object with Lambertian reflectance function, which is illuminated by a single point light source at infinity. Let $B \in \Re^{n \cdot 3}$ be a matrix where each row is the product of the albedo with the inward pointing unit normal for a point on the surface corresponding to a particular pixel in the image. Let $s \in \Re^3$ denote the product of the light source intensity with the unit vector in the direction of the light source. The resulting image $x \in \Re^n$ is then given by

$$x = \max(Bs, 0) \tag{1}$$

The pixels set to zero correspond to the surface points lying in an *attached shadow*. Convexity of the object is assumed to avoid *cast shadows*. When no part of the object is shadowed, x lies in the 3-D subspace L, called the *illumination space*, given by the span of the matrix B, where

$$L = \left\{ x \,\middle|\, x = Bs, \forall s \in \Re^3 \right\} \tag{2}$$

Hence the illumination subspace can be constructed from just three basis images [9].

It was shown in [2] that the set C of all possible images of a convex Lambertian surface, created by varying the direction and strength of an arbitrary number of point light sources at infinity, is defined as follows:

$$C = \left\{ x \,\middle|\, x = \sum_{i=1}^{k} \max(Bs_i, 0), \forall s_i \in \Re^3, \forall k \in Z^+ \right\} \tag{3}$$

and C is a convex cone in $\Re^n$. Furthermore, it was shown in [2] that any image in the cone C can be represented as a convex combination of *extreme rays* given by

$$x_{ij} = \max(Bs_{ij}, 0) \tag{4}$$

where

$$s_{ij} = b_i \times b_j, \quad i \neq j \tag{5}$$

where b_i and b_j are the rows of B It was proved in [2] that the number of shadowing configurations is at most $m(m-1)+2$, where $m \leq n$ is a number of distinct normals. Hence there are at most $m(m-1)$ extreme rays. Since there is a finite number of extreme rays, the convex cone is polyhedral.

The illumination subspace method [2] offers a way to construct the illumination cone. Gather three or more images of the object (with a fixed pose) under varying illumination without shadowing, and use these images to estimate the three-dimensional illumination subspace L by normalizing the images to unit length, and then using singular value decomposition (SVD) to estimate the optimal three-dimensional orthogonal basis B^* in least square sense. It was proved in [2] that B^* is sufficient for determining the subspace L. Then from B^*, the extreme rays defining the illumination cone C can be computed using Eq.4 and 5.

In this paper we use the observations from [2] and the newly introduced *anti-face* method [6] to detect 3-D objects under variable illumination and various classes of geometric transformations. The anti-face method offers an attractive solution, which proceeds by modeling the effects of different illumination conditions in the training set; this automatically voids the illumination effects, allowing fast illumination invariant detection, without having to create a large training set.

Section 2 focuses on applying the anti-face method [6] to the illumination space and illumination cone, and presents novel applications for detection of objects with Lambertian surface under both varying illumination and pose Section 3 introduces an extension of the presented algorithms for ambient light. Section 4 presents the experimental results.

2. Application of Anti-Face Method to Illumination Invariant Detection

Anti-faces [6] is a novel detection method, which works well in case of a rich image collection – for instance, frontal face under a large class of linear transformations, or 3-D objects under different viewpoints. Call the collection of images, which should be detected, a *multi-template* The detection problem is solved by sequentially applying very simple filters (or *detectors*), which act as inner products with a given image (viewed as vector) and satisfy the following conditions:

- The absolute values of their inner product with multi-template images are small.
- They are smooth, which results in the absolute values of their inner product with "random images" being large; this is the characteristic which enables the detectors to separate the multi-template from random images.
- They act in an independent manner, which implies that their false alarms are not correlated; hence, the false alarm rate decreases exponentially in the number of detectors.

The detection process is very simple: the image is classified as a member of the multi-template iff the absolute value of its inner product with each detector is smaller then some (detector specific) threshold. Only images which passed the threshold test imposed by the first detector, are examined by the second detector, etc. This, in turn, leads to a very fast detection algorithm. Typically, $(1+\delta)N$ operations are required to classify an N-pixel image, where $\delta < 0.5$. To achieve invariance to the illumination intensity, the images are normalized to zero mean and unit length.

2.1. Extension of Anti-Faces to Varying Illumination

To extend anti-faces to handle illumination variability we could sample the entire illumination space of the object, and use it as a training set. Obviously this is practically impossible because of the size of the training set. Hence we should find a small number of "basis images" and corresponding detectors such that:

A) After normalization, the different object appearances can all be represented as linear combinations of the basis images, with small combination coeffients.

B) The detectors have small inner products with the basis images. Because of (A), they will also have small inner products with all the object appearances. This will be formalized in Proposition 1.

The following observations [2] support condition (A). Consider a convex object with a Lambertian reflectance function.

- When no part of the object is shadowed, its image lies in the 3-D subspace L given by the span of the matrix B; L can be constructed from three basis images.
- The set of images under an arbitrary number of point light sources at infinity is a convex polyhedral cone in $\Re^n$, which can be expressed as convex combination of extreme rays.

In order to satisfy these conditions, let us first analyze the positive set of the anti-face detector (that is, the set of images accepted by the detector).

Assume that the training set consists of orthonormal vectors. This assumption is feasible, because using SVD we can replace the original training set with the eigenvectors that correspond to the eigenvalues which capture 99% of the energy

Proposition 1. *Let $\{v_i\}_{i=1}^k$ be the orthonormal basis produced by SVD from the normalized training set Let d be an anti-face detector, such that*

$|(d,v_i)| \le \varepsilon_i, \forall i = 1,...,k$ *Then for each* $v = \sum_{i=1}^{k} \alpha_i v_i$, *which satisfies* $\|v\| = 1$, $|(d,v)| \le \sqrt{\sum_{i=1}^{k} \varepsilon_i^2}$.

Proof: omitted for brevity.

From Proposition 1 it follows that if the three basis images for illumination subspace are used as training set for the detector, it will detect the entire illumination subspace if the threshold is properly chosen.

As was previously mentioned, the illumination cone can be represented by linear combination of the vectors x_{ij} (Eq.4) with non-negative coefficients. For the normalized image, the illumination cone is the intersection between the illumination cone for the non-normalized image and the unit sphere in $\Re^n$. Thus from the last observation and Proposition 1 it follow that if the detector is trained on x_{ij} it will detect the entire illumination cone, if the threshold is correctly chosen.

2.2. Illumination Spaces for Various Classes of Geometric Transformations

The following algorithm is designed to detect a convex object with a Lambertian reflectance function under various classes of geometric transformations, when no part of the object is shadowed.

1. Create the training set as follows. For a sample of object positions:
   a. Gather three or more images of the object under varying illumination without shadowing.
   b. Normalize the images to unit length, apply SVD, and take the three eigenvectors that correspond to the largest eigenvalues. Call these the *basis images*.
2. Replace the training set (containing images captured in all the positions), by the eigenvectors $\{v_i\}_{i=1}^{k}$ that correspond to the eigenvalues capturing 99% of the energy.
3. Find anti-face detectors using the new training set.
4. For each detector *d*, choose the threshold as $\sqrt{\sum_{i=1}^{k} \varepsilon_i^2}$, where $|(d,v_i)| = \varepsilon_i$, $i = 1,..,k$.

From Proposition 1, it follows that the positive set of such a detector includes the entire illumination space for all object positions on which the detector was trained.

2.3. Illumination Cone for Fixed Pose

The following algorithm is designed to detect a convex object with a Lambertian reflectance function under fixed pose and arbitrary number of point light sources at infinity. Attached shadows are allowed.

1. Gather three or more images of the object (with a fixed pose) under varying illumination without shadowing.
2. Normalize the images to unit length, and use SVD to estimate the best three-dimensional orthogonal basis B^* in least square sense.
3. From B^* compute the vectors x_{ij} using Eq.4 and. 5
4. Apply SVD on the collection of vectors x_{ij} in order to find the eigenvectors $\{v_i\}_{i=1}^{k}$ that correspond to the eigenvalues capturing 99% of the energy.
5. Find anti-faces using $\{v_i\}_{i=1}^{k}$ as training set and calculated the threshold in the same way as in 2.2.

From Proposition 1, it follows that the positive set of such a detector includes the entire illumination cone.

As mentioned in Section 1, the number of extreme rays is $m(m-1)$ where $m \le n$ is the number of distinct normals, which is usually large, hence the number of extreme rays needed for construction of the illumination cone can run in the millions. Therefore, we use the sampling method from [2], that approximates the cone by directly sampling the space of light source directions rather than generating the samples through Eq.4 and 5.

2.4. Illumination Cones for various classes of geometric transformations

The algorithm from 2.3 can be extended to detect an object under various classes of geometric transformations and arbitrary number of point light sources at infinity. Create the training set by computing the extreme rays for each sample of object positions (Section 2.3) and replace it by the eigenvectors that correspond to the eigenvalues, which capture a 99% of the energy. Find the anti-face detectors and their thresholds in the same way as in 2.3. The positive set of the detectors includes the illumination cone for each object pose.

3. Extension to Ambient Light

The illumination space and illumination cone models assume one or more point light sources at infinity. Such models are unrealistic. When an object is illuminated by a single point light source, the light is diffused by surrounding walls, dust particles, etc., causing the presence of ambient light in the image. Incorporation of ambient light into the previously presented light models extends the 3-D illumination subspace L (Eq. 2) to a 4-D subspace, L_a:

$$L_a = \left\{x \,\middle|\, x = B_a s_a = Bs + \lambda a, \forall s \in \Re^3, \lambda \in \Re_+\right\} \quad (6)$$

where a is an image of object illuminated by ambient light and λ is an ambient light intensity.

In matrix form we have

$$B_a = [B \quad a] \text{ and } s_a = \begin{bmatrix} s \\ \lambda \end{bmatrix} \tag{7}$$

In order to estimate the illumination subspace L we use SVD to find the best three-dimensional orthogonal basis B^* in least square sense. From (7) follows that L_a can be estimated by

$$B_a^* = [B^* \quad a] \tag{8}$$

The following method offers a way to construct the illumination cone for a convex object with Lambertian reflectance function, illuminated by an arbitrary number of point light sources at infinity and ambient light.

1. Gather three or more images $I_1,...,I_p$ of the object under varying illumination without shadowing but including ambient light. Take one image A of an object illuminated only by ambient light. For each $i = 1,..., p$

$$I_i = \widetilde{I}_i + A \tag{9}$$

where $\widetilde{I}_i$ is the image of the object without the ambient component. In order to subtract the ambient component from the image, we project the image on A and subtract the projection:

$$\widetilde{I}_i = I_i - \frac{(I_i, A)}{\|A\|^2} A \qquad i = 1,..., p \tag{10}$$

2. Use $\widetilde{I}_i$ in order to find B^*.
3. From B^* compute the vectors x_{ij} using Eq.4 and Eq.5
4. Use $\{x_{ij}\} \cup A$ for constructing the illumination cone extended for ambient light.

The proposed method can be incorporated into all algorithms described in Section 2.

4. Experimental Results

In all experiments, we used the multi-template created from 30x30 gray-level images. For nearly all light directions, there were no cast shadows present, and the object was correctly detected. It is also possible to follow [4], and directly remedy the cast shadows. We have chosen image plane rotations for training and testing the algorithms described in Section 2.2 and 2.4.

Ten images of the tiger doll were captured under varying illuminations without shadowing (Figure 1a). The object was illuminated by a single light source, but because of diffusion from the surroundings the ambient light component is present in all images. We also captured one image under ambient light only (Figure 1b). Using the algorithm from Section 2 we found the three basis images that span the illumination subspace L (Figure 1c). Figure 2 presents the results of the detection algorithm, under arbitrary rotations and various illuminations without shadowing (Section 2.2). Ten detectors were sufficient to recover the doll without false alarms. The anti-face method [6] that was trained on the image of the doll subject to arbitrary rotations and illuminated by ambient light alone failed to detect the object in the scenes depicted in Figure 2.

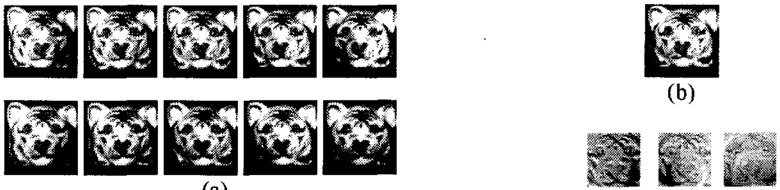
The following experiment was designed to test the algorithm for detecting an object under fixed pose, illuminated by an arbitrary number of point light sources at infinity. Attached shadows were allowed. We took the same basis images (Figure 1c) as before, and used the sample method [2] to approximate a cone. It was empirically shown in [2] that the cone is flat (i.e., its elements lie near a low dimensional linear subspace), and that the subsampled cone provides an approximation that results in good recognition performance. In our experiment we created about 60 images, such that the corresponding light source directions s_{ij} were more or less uniformly distributed on the illumination sphere [4]. Figure 3 demonstrates the results of the detection of the tiger doll in real images under various illuminations. Eight to ten anti-face detectors were used to detect all the instances of tiger with no false alarms.

Since it is very difficult to simulate the light conditions that result in the images with significant attached shadows, we tested the algorithm on 200 random samples from the illumination cone of the tiger with one and two light sources. The images where artificially created using the method described in [2]. All 200 samples were recognized as the tiger. Figure 4 shows some of the images from this test set.

The last experiment was designed to test the algorithm for detecting illumination cones for image plane rotations with a range of 45 degrees (Section 2.4). We created the extreme rays for each rotation angle in the manner described in the previous experiment. Eight sets of anti-face detectors were created, each one for a 45 degree range, thus covering 360 degrees. Figure 5a contains the image of the tiger rotated by 180 degrees. Figure 5b presents the image rotated by 60 degrees. In these tests, ten anti-face detectors sufficed to detect the tiger without false alarms.

5. Conclusions

We have presented a novel algorithm for detection of convex objects with Lambertian surface under various classes of geometric transformations and variable illumination, which accounts for attached shadows. The key element of our approach was to model the effects of different illumination conditions that can be learned from a small set of images [2] in the training set of the anti-face



(a) (b) (c)

Figure 1: Initial images for estimation of illumination space. (a) images illuminated by a single light source and ambient light; (b) image illuminated by ambient light only. (c) basis images that span the illumination subspace for the tiger doll.

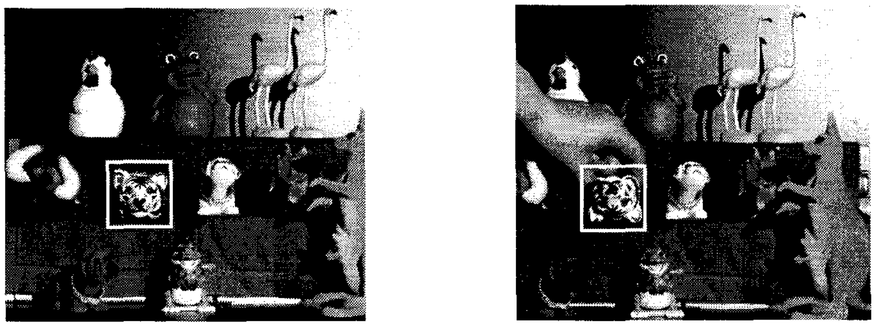
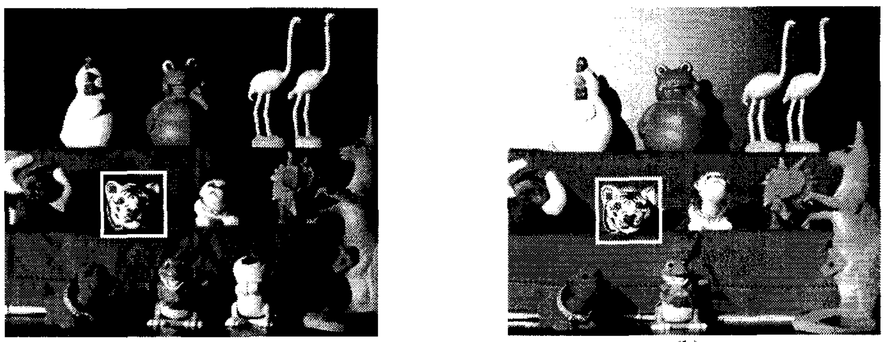
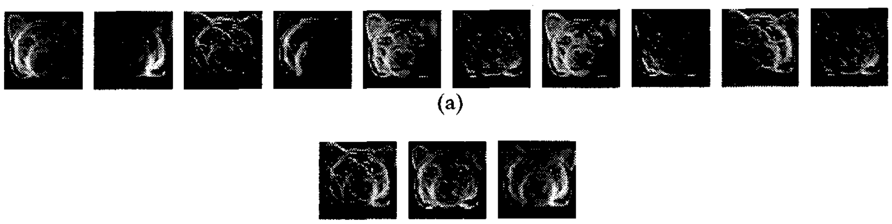


Figure 2: Results of detection of the tiger doll, subject to image plane rotations and various illuminations without shadowing. The scene was illuminated by point light source and ambient light.



(a) (b)

Figure 3: Detection results in real images; (a) one light source and ambient light; (b) two light sources and ambient light.



(a)

(b)

Figure 4: Random samples from illumination cone of tiger: (a) – with one and (b) – with two light sources.

(a) (b)

Figure 5: Detection results for tiger under 45-degree range of rotations and illumination by a single point light sources and ambient light. (a) tiger doll is rotated by 180 degrees; note very strong shadowing effect; (b) image rotated by 60 degrees.

detectors; this automatically voids the illumination effects, allowing fast illumination invariant detection. Furthermore we have extended this method to incorporate ambient light. The method was successfully applied to detect an object under variable illumination and rotations in real images with complicated background and simulated images with significant attached shadows.

The experiments described above were limited to plane rotation of an object. In further research we plan to test the presented approach on the images containing a 3-D object under various illuminations and pose.

6. Acknowledgement

The authors would like to thank Daniel Osadchy for assisting with the image capture, and for many useful discussions.

7. References

[1] P. N. Belhumeur, P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", *IEEE Trans. on Patt. Anal. and Mach. Intelligence,* 19(7):711-720, 1997.

[2] P. N. Belhumeur and D. J.Kriegman,"What is the set of images of an object under all possible lighting conditions?" *Int.. J. Comp. Vision,* 28(3): 245-260, 1998.

[3] A.S. Georghiades, P. N. Belhumeur, and D. J. Kreigman, "From few to many: Generative Models for Recognition under variable pose and illumination," *Conf. On Automatic Face and Gesture Recognition, 2000.*

[4] A.S. Georghiades, D. J. Kreigman, and P. N. Belhumeur, "Illumination cones for recognition under variable lighting: Faces", *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.,* 1998.

[5] H. Hayakawa, "Photometric stereo under a light-source with arbitrary motion", *JOSA-A,* 11(11):3079-3089, Nov,1994.

[6] D. Keren, M. Osadchy, and C. Gotsman. "Anti-Faces: A novel, fast method for image detection," To appear in *IEEE Trans. on Patt. Anal. and Mach. Intelligence.*

[7] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation", *Trans. on Pattern Anal. and Mach. Intelligence,* 19(7):696-710, 1997.

[8] H. Murase and S. Nayar, "Visual learning and recognition of 3-D objects from appearance", *Int. J. Comp. Vision,* 14(5-24), 1995.

[9] S. Nayar and H. Murase, "Dimensionality of illumination in appearance matching", *IEEE Conf. on Robotics and Automation,* 1996.

[10] E. Osuna, R.Freund and F. Girosi. "Training support vector machines: An application to face detection", *IEEE Conf. on Comp. Vision and Patt. Recog.,* 1997.

[11] C. P. Papageorgiou, M.Oren, and T. Poggio, "A general framework for object detection", *Int. Conf on Comp. Vision,* pp. 555-562, 1998.

[12] M. Pontil and A. Verri , "Support vector machines for 3D object recognition", *IEEE Trans. on Patt. Anal. and Mach. Intelligence,* 20(6):637-646, 1998.

[13] T. Riklin-Ravin and A. Shashua, " The quotient image: class based re-rendering and recognition with varying illuminations", *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.,* pp. 566-571, 1999.

[14] D. Roobaert and M.M. Van Hulle, "View-based 3D object recognition with support vector machines", *IEEE Int. Workshop on Neural Networks for Signal Processing,* USA 1999, pp. 77-84.

[15] H. A. Roweley, S. Baluja, and T.Kanade. "Neural network-based face detection", *IEEE Trans. on Patt. Anal. and Mach. Intelligence,* 20(1):23-38, 1998.

[16] H. A. Roweley, S. Baluja, and T.Kanade "Rotation invariant neural network-based face detection", *IEEE Conf. on Comp. Vision and Patt. Recog.,* 1998.

[17] A.Shashua, *Geometry and Photometry in 3D Visual Recognition,* PhD thesis, MIT, 1992.

[18] K.K. Sung and T.Poggio, "Example based learning human face detection", *IEEE Trans. on Patt. Anal. and Mach. Intelligence,* 20(1):39-51, 1998.

[19] M. Turk and A. Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience,* 3(1):71-86,1991.

[20] M. Turk and A. Pentland, "Face Recognition using eigenfaces," *Proc. of the Int'l Conf.on Comp. Vision and Patt. Recog.,* 1991, pp. 586-591.