Using Stereo Matching for 2-D Face Recognition Across Pose

Carlos D. Castillo David W. Jacobs
Computer Science Department, University of Maryland, College Park
{carlos, djacobs}@cs.umd.edu

Abstract

We propose using stereo matching for 2-D face recognition across pose. We match one 2-D query image to one 2-D gallery image without performing 3-D reconstruction. Then the cost of this matching is used to evaluate the similarity of the two images. We show that this cost is robust to pose variations. To illustrate this idea we built a face recognition system on top of a dynamic programming stereo matching algorithm. The method works well even when the epipolar lines we use do not exactly fit the viewpoints. We have tested our approach on the PIE dataset. In all the experiments, our method demonstrates effective performance compared with other algorithms.

1. Introduction

Face recognition is a fundamental problem in computer vision. There has been a lot of progress in the case of images taken under constant pose [15]. There are several approaches to handling pose variation [11, 7, 9]. However, there is still a lot of room for improvement. Progress would be important for applications in surveillance, security, the analysis of personal photos and other domains in which we cannot control the position of subjects relative to the camera.

Correspondence seems quite important to produce meaningful image comparisons. Existing systems often align the eyes or a few other features, using translation, similarity transformations, or perhaps affine transformations. However, when the pose varies these can still result in fairly significant misalignments in other parts of the face, as we will demonstrate.

In many real applications of face recognition, the epipolar geometry of the cameras is approximately known even when the distance between cameras is unknown. In particular, when we take upright pictures of people, as in personal photos, the epipolar lines are approximately horizontal. Therefore,we propose using stereo matching to produce a measure of the similarity of two faces (in unknown poses). We show that the matching cost is robust to horizontal pose variations. Note that we are not interested in performing 3-D reconstruction. We show that the method works well even when the epipolar lines we use are not accurate due to a significant difference in height in the two cameras.

The rest of the paper is organized as follows. Sec. 2 discusses related work. Sec. 3 analyzes the use of stereo matching algorithms for recognition across pose. Sec. 4 presents the details of our face recognition method and Sec. 5 presents and analyzes all experiments. Sec. 6 concludes.

2. Related Work

Zhao et al. [15] review the vast literature on face recognition. Although the bulk of this work assumes fixed pose, there have been a number of approaches that do address the problem of pose variations. Many of these methods use some 3-D knowledge of faces to compensate for pose.

Blanz and Vetter [3] use laser scans of 200 subjects to build a general morphable model of three dimensional faces. Then, with the aid of manually selected features, they fit this model to images. The parameters of the fit to two different images can be compared to perform recognition. In their experiments they show strong results for a subset of the poses in the PIE database.

In Romdhani et al. [11] shape and texture parameters of a 3-D morphable model are recovered from a single image. They present exhaustive results of experiments with pose variations for the PIE dataset and show strong results (the best results we're aware of with pose variation).

Basri and Jacobs [1] use a 3D model to generate a low dimensional subspace containing all the images that an object can produce under lighting variation. Pose is determined using manually selected point features.

In Georghiades et al. [6] a 3-D head model is computed for each person using a gallery containing a number of images per subject taken with controlled illumination. Pose variation is handled by sampling the set of possible poses, and building a 2-D model for each one. They evaluate their method using the Yale Face Database B. However, it is not clear how such a method might perform using arbitrary galleries and probes.

In Gross et al. [7] two appearance-based algorithms for

1-4244-1180-7/07/$25.00 ©2007 IEEE

face recognition across pose and illumination are presented. One of them is called eigen light-fields. At the core of the method is the *plenoptic function* or light field. To use this concept, all of the pixels of the various images are used to estimate the (eigen) light-field of the object. They evaluate their results using the CMU PIE dataset [12]. In its assumptions, (recognizing faces across general unknown poses), this method is the most similar to ours. However our approach is simpler and our results are better.

The other method presented in Gross et al. [7] is called Bayesian Face Subregions (BFS). The algorithm models the appearance changes of the different face regions in a probabilistic framework. Using probability distributions for similarity values of face subregions, the method computes the likelihood of probe and gallery images coming from the same subject. The method is particularly useful when the probe pose is known.

Beymer and Poggio [2] generate 2-D virtual views from a single image per person using prior knowledge of the object class (in particular symmetry and prototypical objects of the same class) using optical flow. Once the virtual view has been generated the images are compared. The method is similar to ours in the sense that it is decidedly 2-D and that it stresses the importance of having good correspondences for face recognition across pose.

Many other approaches compensate for some 2-D deformations in matching, which may partially compensate for the effects of pose. A notable example is the work of Wiskott et al [14]. They developed a method called Elastic Bunch Graph Matching (EBGM) that is a simplified implementation of dynamic link architecture methods based on a neural network and a geometric measure.

3. Analysis of Stereo Matching for Face Recognition

Most work in image-based recognition aligns regions to be matched with a low-dimensional transformation, such as translation, or a similarity or affine transformation. Instead, we use stereo matching. When we enforce the ordering constraint, this allows for arbitrary, one-to-one continuous transformations between images, along with possible occlusions, while maintaining an epipolar constraint. In this section we show that the greater generality afforded by stereo matching may be necessary for face recognition, and that stereo matching will not be too sensitive to noise in determining the epipolar lines.

We illustrate this using a very simplified model of faces, in which we calculate the disparity maps that will correctly match two images.

1. We model the face as a cylinder. Perturbations to this model, such as adding a nose, can be handled fairly easily.

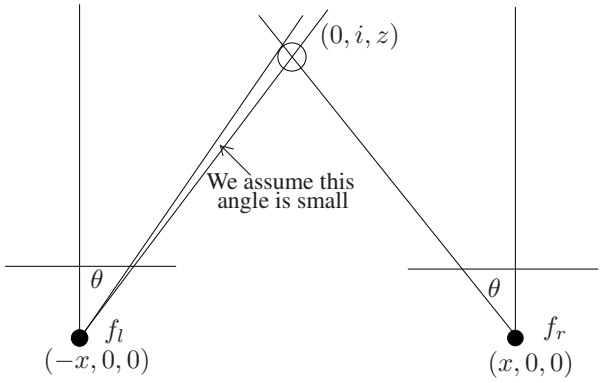


Figure 1. Our very simplified model of faces.

2. We assume the face is viewed by two cameras with image planes that are rectified to be perpendicular to the z axis and that the cylinder axis is the y axis. This is roughly the situation when an upright person photographs another upright person. For simplicity, we will assume that the cylinder lies on the z axis, that the camera focal points lie on the x axis at points symmetric about the z axis (see Figure 1). We call the left and right focal points f_l and f_r respectively.

3. We assume that the distance from the camera to the person is much bigger than the radius of the cylinder that represents the person. Specifically, we assume that vectors from the camera focal point to any location on a horizontal cross section of the cylinder have the same direction. If we imagine that the cylinder (face) has a radius of three inches, and the distance from the camera to the face is 8 feet, we can calculate that a vector from the focal point to the center of a cross-section of the cylinder will be within 5.5 degrees of a vector to any point on the cylinder cross section, so this approximation is not too bad.

These assumptions simplify our presentation, which could be readily extended to other settings.

We will analyze disparities on the $y = 0$ plane. Given these assumptions, each camera will see half of a circular cross-section. They will not see exactly the same half-circle, however, as there will be some occlusion. Without loss of generality assume the radius of the circle is 1. We will denote the angle between the z axis and a vector from f_l to the cylinder by θ. The corresponding angle for the right camera will then be $-\theta$. Define l_1 and l_2 to be two points on the circle, such that the tangent lines to the circle at l_1 and l_2 pass through f_l. That is, l_1 and l_2 are the first and last points on the circle that are visible in the left image. Define L to be the line connecting l_1 and l_2. We can similarly define r_1 and r_2 for the right image. So, for example, the region of the circle between r_1 and l_2 is visible in both images.

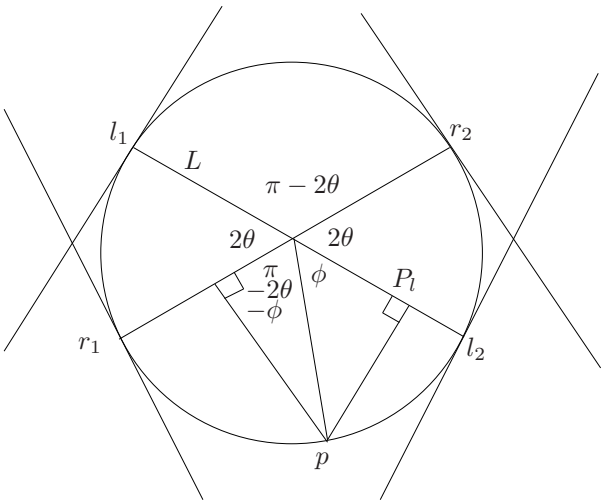


Figure 2. The circle paramerized by the angle ϕ.

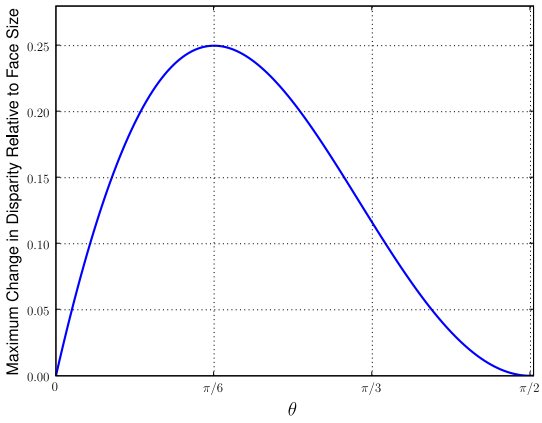


Figure 3. Change in disparity relative to the size of the face as a function of θ.

Note that every line connecting f_l to L intersects the circle in a single point that will be visible in the left camera. So one way to determine the image of the circle in the left camera is to project the visible half-circle onto L using these lines, and then to consider how L is projected onto the left camera. Because we assume the cylinder is small relative to its distance to the camera, we can approximate the projection of L into the left camera using scaled-orthographic projection. Without loss of generality we can normalize the left image so that the width of the circle's projection is 1 (this is in image units, which may differ from 3D units), and the x coordinate of the image of l_1 is 0. This is illustrated in Figure 2.

We can parameterize points on the circle by the angle ϕ, which we take relative to l_2 (see Figure 2). Consider some such point p. We can determine the location of p in the left image, by considering the line through p and f_l. The point where this line intersects L, call it P_l, will appear in the same image location as p. Define the distance from P_l to l_1 to be $d(l_1, P_l)$. Then the x coordinate of p in the left image is $d(l1, P_l)/2 = (1 + \cos\phi)/2$. Similarly, its position in the right image will be $(1 - \cos(\pi - 2\theta - \phi))/2$. If we define the disparity, d, in a matched point to be its x coordinate in the left image minus the x coordinate in the right; we get:

$$d = (\cos\phi + \cos(\pi - 2\theta - \phi))/2 \tag{1}$$

It is straightforward to show that disparity is minimized by $\phi = 0$ or $\phi = \pi - 2\theta$, which are the furthest points visible in both cameras, and maximized by $\phi = (\pi - 2\theta)/2$, which corresponds to the point closest to the cameras.

We are interested in the variation between the minimum and maximum disparity values, Δd. We have:

$$\Delta d = \cos\left(\frac{\pi}{2} - \theta\right) - \frac{1}{2} - \frac{\cos(\pi - 2\theta)}{2} \tag{2}$$

This is maximized for $\theta = \pi/6$, when $\Delta d = 1/4$. Figure 3 shows how the maximum change in disparity varies with θ. In Figure 3, we can see that for a large range of θ, disparity changes quite a bit within the image.

From this analysis, we can see that for a cylinder, disparity in an image can vary by as much as 1/4 of the apparent width of the cylinder, and frequently varies substantially. These variations in disparity cannot be accounted for by aligning the images with a linear transformation, since linear transformations can only create linear disparity maps. In scenarios such as the one described here, because of the symmetry of the viewing conditions, we can demonstrate that the optimal linear transformation to align the two images will simply be the identity transformation, which does not account for any of these variations in disparity. Note that the amount of disparity is independent of the distance from the cameras to the face, because we measure disparity relative to the apparent size of the face.

Ideally, one should determine the epipolar geometry prior to matching two faces. However, in many cases, images result from an upright photographer taking a picture of an upright subject. This results in epipolar lines that are approximately horizontal. If we align the eyes in two photographs, this will align corresponding horizontal epipolar lines. However, error will result when epipolar lines are not purely horizontal. To get a sense of the possible magnitude of this error, we analyze a simple example.

Consider the case in which we take two pictures of a face that is five feet high, at a distance of eight feet. But suppose that the disparity is vertical instead of horizontal, because one photograph is taken from a height of five feet, and the second is taken from a height of six feet. Vertical disparity will be zero at the eyes, which are aligned, and will be maximized at the point that is closest to the cameras,

the tip of the nose. If we assume that the nose is about one inch long, then using similar triangles we can determine that it appears at the same image location as a point 1/8 of an inch below the nose, in the second image. For a face that is six inches long, the vertical disparity will therefore be about 2% of the height of the face in the image. This error is small compared to the variations of up to 25% in horizontal disparity that can arise in the situation we analyze above. Of course, this is just an illustrative example; the error introduced by mis-estimation of the epipolar lines will depend in practice on the viewing conditions typical in a specific application. Our example simply makes the point that in some common settings, this error will be quite small, while stereo matching can compensate for correspondence errors that will be large.

4. Stereo Matching and Face Recognition

We require an efficient stereo algorithm appropriate for wide baseline matching of faces. A number of methods might be suitable. We have used Criminisi et al. [5][1] which has been developed for video conferencing applications and so seems to fit our needs. It is not obvious that it will work for the large changes in viewpoint that can occur in face recognition, but we will show that it does.

It is important to stress that we are relatively unaffected by some of the difficulties that make it hard to avoid artifacts in stereo reconstruction. For example, when many matches have similar costs, matching is ambiguous. Selecting the right match is difficult, but important for good reconstructions. However, since we only use the cost of a matching, selecting the right matching is unimportant to us in this case. Also, errors in small regions, such as at occluding boundaries, can produce bad artifacts in reconstructions, but that is not a problem for our method as long as they don't affect the cost too much.

The core of the stereo method calculates a matching between two scanlines (rows of each face). We use the 4 planes, 4 transitions stereo matching algorithm described in [5]. This is a dynamic programming stereo matching algorithm that is fast and performs well when compared to other methods. The benefit of this formulation is having more control than traditional one plane models [4].

The method defines 4 planes (or matrices) called C_{Lo}, C_{Lm}, C_{Ro} and C_{Rm}, which capture matching and occlusions in the left and right images. The planes naturally define the persistence of states. By setting the state transition costs adequately many state transitions can be favored or biased against, for example long runs of occlusions can be favored over many short runs by setting a high cost for entering or leaving an occluded state.

[1]We also tried the method described in Cox et al. [4] and found the method to be a bit faster but less accurate than the method described in Criminisi et al. [5].

The elements of the cost matrix are initialized to $+\infty$ everywhere except in the right occluded plane where:

$$C_{Ro}[i,0] = i\alpha \quad \forall i = 0 \dots W-1 \tag{3}$$

The forward step of the 4-state DP computes the four cumulative cost matrices according to the following recursion:

$$C_{Lo}[l,r] = \min \begin{cases} C_{Lo}[l,r-1] + \alpha \\ C_{Lm}[l,r-1] + \beta \\ C_{Rm}[l,r-1] + \beta \end{cases} \tag{4}$$

$$C_{Lm}[l,r] = M(l,r) + \min \begin{cases} C_{Lo}[l,r-1] + \beta' \\ C_{Lm}]l,r-1] + \gamma \\ C_{Rm}[l,r-1] \\ C_{Ro}[l,r-1] + \beta' \end{cases} \tag{5}$$

where $M(l,r)$ is the cost of matching the lth pixel in the left scanline with the rth pixel in the right scanline. α, β, β' and γ are parameters that can be set experimentally. C_{Ro} and C_{Rm} are symmetric. Our experiments show that the method is rather insensitive to these parameters and all experiments shown here are run with $\alpha = 0.5$, $\beta = \beta' = 1.0$ and $\gamma = 0.25$ as recommended in [5]. $M(l,r)$ is a fast approximation to the normalized cross correlation of a 3×7 window around the points (l,s) and (r,s) of the images, where s is the current scanline.

The cost of matching the two scan lines l_1 and l_2, denoted $\mathrm{cost}(l_1,l_2)$, is: $C_{Ro}[l-1,r-1]$. The optimal matching solution will be a sequence of symbols in the alphabet: $\{C_{Lo}, C_{Lm}, C_{Ro}, C_{Rm}\}$ which can be obtained by following a backward step. But we have no use for the optimal matching, we only use its cost.

Let $\mathrm{cost}(I_1,I_2)$ define the cost of matching the rows of I_1 with the rows of I_2, as defined above:

$$\mathrm{cost}(I_1,I_2) = \sum_{i=1}^{n} \mathrm{cost}(I_{1,i}, I_{2,i}) \tag{6}$$

where $I_{1,i}$ is the i-th scan line (row) of image 1. Given two images with unknown pose, I_1 and I_2, we define the similarity of the two images as:

$$\mathrm{similarity}(I_1,I_2) = \min \begin{cases} \mathrm{cost}(I_1,I_2) \\ \mathrm{cost}(I_2,I_1) \\ \mathrm{cost}(\mathrm{flip}(I_1),I_2) \\ \mathrm{cost}(I_2,\mathrm{flip}(I_1)) \end{cases} \tag{7}$$

since we do not know which image is left and which image is right we have to try both options, one of them will be the true cost, the other cost will be noise and should be ignored. Additionally, *flip* produces a left-right reflection of the image. *flip* is helpful when two views see mainly different

sides of the face. In this case, a truly correct correspondence would mark most of the face as occluded. However, since faces are approximately vertically symmetric, *flip* approximates a rotation about the y axis that creates a virtual view so that the same side of the face is visible in both images. For example, if we viewed a face in left and right profile, there would be no points on the face visible in both images, but flipping one image would still allow us to produce a good match.

Finally, we perform recognition simply by matching a probe image to the most similar image in the gallery. For the method to work well all the images in the gallery have to be in the same pose.

5. Experiments

The experimental evaluation is separated into four parts. All the experiments were run with the CMU PIE database [12]. The CMU PIE database consists of 13 poses of which 9 have approximately the same camera altitude (poses: c34, c14, c11, c29, c27, c05, c37, c25 and c22). Three other poses that have a significantly higher camera altitude (poses: c31, c09 and c02) and there is one last pose that has a significantly lower camera altitude (pose c07). We say that two poses have aligned epipolar lines if they are both from the set: {c34, c14, c11, c29, c27, c05, c37, c25, c22}. And we say that two poses have misaligned epipolar lines if one comes from the set {c34, c14, c11, c29, c27, c05, c37, c25, c22} and the other comes from the set {c31, c09, c07, c02}.

The faces were cropped to a size of 40x48 pixels showing only the face. The images were aligned using manually selected feature points with a similarity transformation.

To be able to compare results with [7, 9] we needed to use a subset of 34 people because they use 34 people for training and the remaining 34 for testing. We don't require training, but we're interested in comparing the methods in equal conditions so we tested on individuals 35-68 from the PIE database. To compare with [11] we used the whole set of 68 people.

First, we evaluate our method with pose variation but fixed lighting. This is done in two separate experiments, one to compare with [7, 9] and the other to compare with [11]. Then to illustrate that our method works in more realistic situations we evaluated simultaneous variation in pose and illumination. This too is done in two separate experiments, one to compare with [7, 9] and one to compare with [11].

5.1. Pose Variation: Comparison with Gross et al.

We conducted an experiment to compare our method with four others. We compared with two variants of eigen light-fields[7], eigenfaces[13] and FaceIt as described in [7, 9]. FaceIt[2] is a commercial face recognition system from Identix which finished top overall in the Face Recognition Vendor Test 2000. Eigenfaces is a common benchmark algorithm for face recognition. Finally, eigen light-fields is a state of the art method for face recognition across pose variation.

Table 1. A comparison of our stereo matching distance with other methods accross pose.

34 Faces

Method	Accuracy
Eigenfaces [7, 9]	16.6%
FaceIt [7, 9]	24.3%
Eigen light-fields (3-point norm.) [7, 9]	52.5%
Eigen light-fields (Multi-point norm.) [7, 9]	66.3%
Stereo Matching Distance	82.0%

68 Faces

Method	Accuracy
Stereo Matching Distance	73.5%
LiST (Romdhani et al. [11])	74.3%

In this experiment we selected each gallery pose as one of the 13 PIE poses and the probe pose as one of the remaining 12 poses, for a total of 156 gallery-probe pairs. We evaluated the accuracy of our method in this setting and compared to the results in [7, 9]. Table 1 summarizes the average recognition rates. Figure 4 shows several cross-sections of the results with different fixed gallery poses.

The fact that the method performs solidly both when the epipolar lines fit (with an average of 87.4%) and when they don't (with an average of 79.0%) and overall (with an average of 82.0% as reported in Table 1) shows the generality of our method.

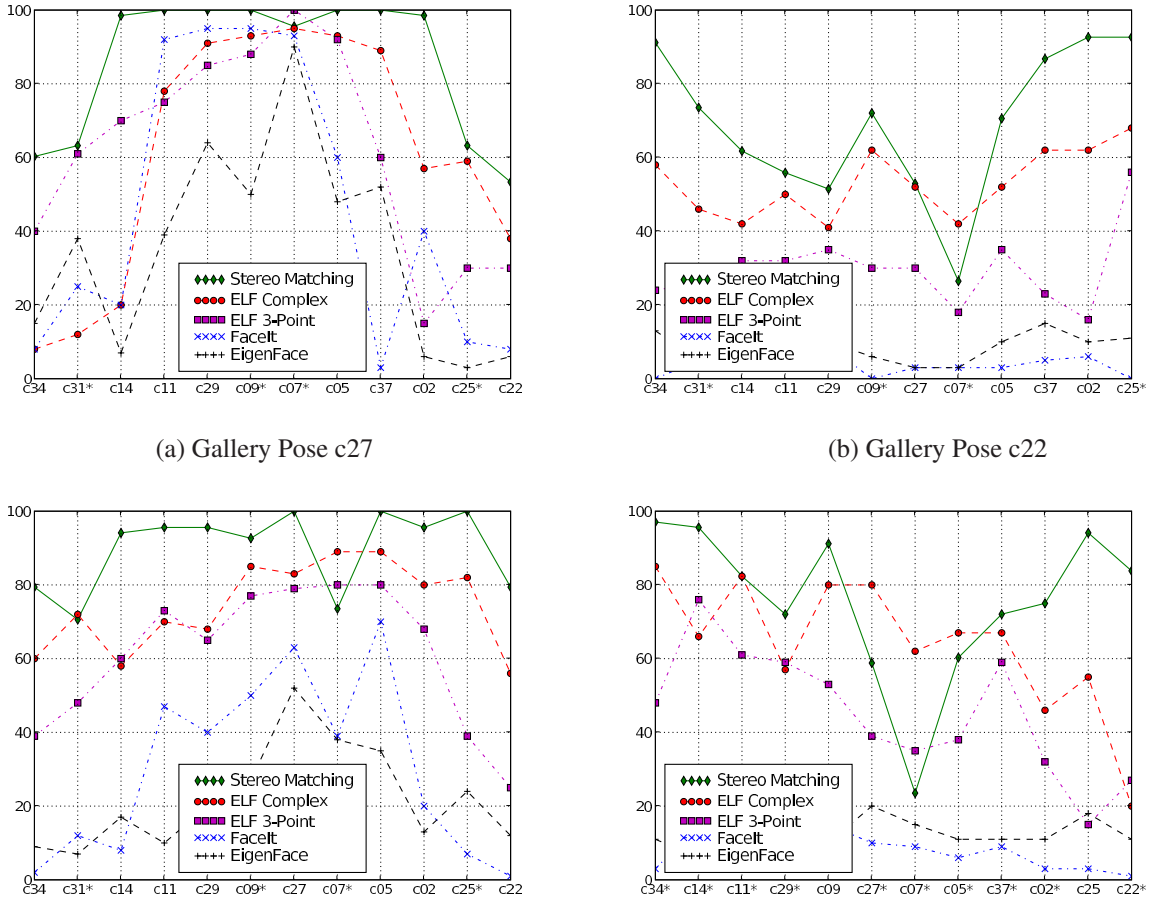
In this experiment we observe that in all gallery poses our method especially outperforms all the other methods for the extreme probe poses (c34, c31, c14, c02, c25 and c22).

5.2. Pose Variation: Comparison with Romdhani et al.

We also compared our results with the ones presented in Romdhani et al[11]. These results are, to our knowledge, the best reported on the whole PIE database for pose variation. In this work all 68 images were used, so for this part we report our results using all 68 faces. Table 1 summarizes the results of this experiment.

The global average for their method is 74.3%, the global average for our method is 73.5%. For the subset of poses in which the epipolar lines fit perfectly our average performance is 80.3%, while theirs is 71.6%. We consider the case where all epipolar lines fit to be our best possible scenario.

[2] Version 2.5.0.17 of the FaceIt recognition engine was used.



(a) Gallery Pose c27

(b) Gallery Pose c22

(c) Gallery Pose c37

(d) Gallery Pose c31

Figure 4. Cross-sections with fixed gallery pose for the results presented in Table 1. Probe poses marked with * have a vertical misalignment of about 10 degrees with the corresponding gallery pose.

Our method runs about 40 times faster than the method presented in [11], and is much simpler. Detailed results are presented in Table 2.

5.3. Variation in Pose and Illumination: Comparison with Gross et al.

We then evaluated the performance of the method across pose and illumination. One of the objectives of this experiment is to verify that the good performance obtained when there is variation in pose (the previous experiments) are not an artifact of the (constant) illumination condition.

In this section we compare our method to Bayesian Face Subregions (BFS) [7] in the case of simultaneous variation of pose and illumination. For this experiment, the gallery is frontal pose and illumination. For each probe pose, the accuracy is determined by averaging the results for all 21 different illumination conditions. The results of this comparison are presented in Figure 5. We observe that our algorithm strictly dominates BFS over all probe poses.

For lighting invariance they use [8] which computes the reflectance and illumination fields from real images using some simplifications, while we simply use an approximation to normalized correlation.

5.4. Variation in Pose and Illumination: Comparison with Romdhani et al.

We performed our experiments in such a way that we can compare with [3] and [11]. For this experiment we used images of the faces of 68 individuals viewed from 3 poses (front: c27, side: c5 and profile: c22) and illuminated from 21 different directions. We used light number 12 for the gallery illumination to be able to compare our results with [11]. They select that lighting because "...the fitting is generally fair at that condition". Our results are presented in

Table 2. Confusion matrix for pose variation. The diagonals are not included in any average. The table layout is the same as [11] and [10].

azimuth	-66	-47	-46	-32	-17	0	0	0	16	31	44	44	62	
altitude	3	13	2	2	2	15	2	1.9	2	2	2	13	3	
Probe Pose	c34	c31	c14	c11	c29	c09	c27	c07	c05	c37	c25	c02	c22	avg
Gallery Pose														
c34	-	91	88	66	59	62	47	16	47	57	65	60	90	62
c31	96	-	94	78	62	82	46	19	56	65	90	66	76	69
c14	85	85	-	100	99	66	90	60	85	82	57	81	47	78
c11	71	75	100	-	100	72	96	81	82	88	47	84	49	79
c29	60	68	100	100	-	94	100	84	97	93	46	84	43	81
c09	66	91	71	85	94	-	99	54	97	90	90	82	68	82
c27	57	56	97	100	100	100	-	90	100	99	50	93	49	83
c07	10	18	60	74	82	56	81	-	79	62	16	47	13	50
c05	43	47	91	93	97	93	100	85	-	100	69	100	62	82
c37	71	63	87	90	90	88	97	65	100	-	87	100	76	85
c25	75	91	57	44	49	91	51	19	72	82	-	93	94	68
c02	68	60	87	85	87	71	96	53	99	100	82	-	85	81
c22	85	65	50	38	40	57	38	16	53	69	88	85	-	57

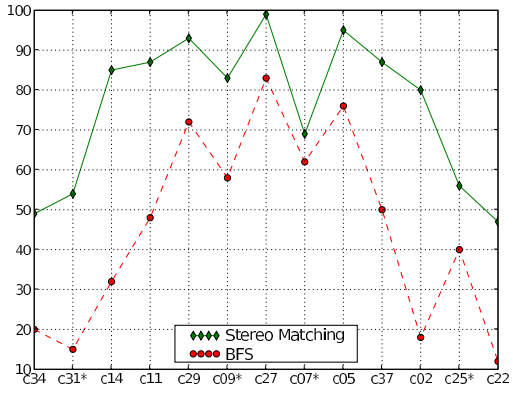


Figure 5. A comparison of out method with BFS. Gallery pose is frontal (c27) probe poses are as indicated in the x axis, we report the average over the 21 illuminations.

Table 3. We do not expect our results to be as good as those of [11], because our algorithm only accounts for lighting variation by using a fast approximation to normalized cross correlation as described in Criminisi et al. [5], while [11] has a 3-D model and performs an optimization to solve for the lighting that best matches the model to the image.

Our stereo matching method degenerates into an approximation to normalized correlation over small windows when there is no change in pose. Our method performs better than Romdhani et al. [11] when there is no pose change (gallery probe combinations: F-F, S-S and P-P). It is surprising that our method works better than theirs in this case because we're using a simple illumination insensitive image comparison technique and they perform an optimization to solve for lighting.

Table 3. Accuracy percentage with pose and illumination variation. Three galleries and three probes were used. F: Frontal, S: Side, P: Profile. Light 0 is ambient lighting. The table layout is the same as [11].

	F Gallery			S Gallery			P Gallery		
light	F	S	P	F	S	P	F	S	P
0	100	100	54	99	100	65	38	41	100
1	67	69	38	61	88	15	28	25	75
2	79	76	34	76	87	25	29	19	76
3	84	82	32	85	94	16	34	28	66
4	93	96	34	93	97	26	34	24	75
5	100	97	35	96	100	46	35	19	99
6	99	92	36	83	99	41	22	19	99
7	100	100	39	100	100	71	29	31	100
8	100	100	41	100	100	71	35	46	100
9	91	91	42	99	90	31	32	16	75
10	100	100	43	100	100	65	38	41	100
11	100	100	43	100	100	65	40	50	100
12	-	100	43	100	-	57	38	44	-
13	100	99	43	100	100	50	43	56	100
14	100	94	42	100	100	38	46	54	100
15	100	78	40	97	99	32	32	35	100
16	99	66	39	91	96	24	37	37	99
17	90	88	38	91	96	16	35	22	60
18	96	97	38	96	99	31	35	21	91
19	100	100	38	97	100	59	35	38	100
20	100	100	39	100	100	57	35	51	100
21	100	97	39	100	100	44	36	54	100
avg	95	92	38	94	98	43	35	35	92

For this experiment our global average is 74% while the global average of Romdhani et al. [11] is 81%, which is considerably better. When there is pose change but no epipolar misalignment and no light change, we perform bet-

ter. When there is light change and no pose change we perform better. This leads us to think that the difference is in the interaction between pose variation and illumination variation.

6. Conclusion

We proposed a method to recognize faces across pose. Compared to existing methods ours is very simple and performs very well. There still is room for improvement in our method, in the sense that a rich variety of more sophisticated strategies can be pursued at each step. As it is, our method illustrates a general, simple, viable alternative to 3-D methods for face recognition across pose [11, 3].

Our method has several interesting properties: (1) it is robust to vertical disparities even when we don't account for them, (2) it degrades gracefully with changes of light and (3) provided that good correspondences are obtained, the method degenerates into normalized correlation over small windows when there is no variation in pose.

While evaluation with the PIE dataset shows that our method performs very well, evaluation of the method in a larger, less controlled database remains to be done.

For the case of only pose variation our results are better than the results of Gross et al. [7, 9]. In the case of only pose variation, our results are comparable to the results of Romdhani et al. [11] but our method has the benefit of being simpler and faster.

For the case of simultaneous pose and illumination variation, our results are better than the results of [7] but worse than the results of [11]. This is not surprising since we don't do anything special to account for variation in illumination.

Finally, we have presented a general method that has a simple, well studied way to account for pose variation and a simple, well studied way to account for light variation and our experiments show that the method works well.

Acknowledgements: Carlos Castillo is supported by a Horvitz Assistantship. We would like to thank Olga Veksler for her advice about stereo.

References

[1] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):218–233, 2003.

[2] D. Beymer and T. Poggio. Face recognition from one example view. Technical Report AIM-1536, , 1995.

[3] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1063–1074, 2003.

[4] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs. A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3):542–567, 1996.

[5] A. Criminisi, J. Shotton, A. Blake, C. Rother, and P. H. S. Torr. Efficient dense stereo with occlusion by four-state dynamic programming. *Intl. Journal of Computer Vision*, 2006.

[6] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001.

[7] R. Gross, S. Baker, I. Matthews, and T. Kanade. Face recognition across pose and illumination. In S. Z. Li and A. K. Jain, editors, *Handbook of Face Recognition*. Springer-Verlag, June 2004.

[8] R. Gross and V. Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*. Springer, June 2003.

[9] R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light-fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(4):449 – 465, April 2004.

[10] R. Gross, J. Shi, and J. Cohn. Quo vadis face recognition? In *Third Workshop on Empirical Evaluation Methods in Computer Vision*, December 2001.

[11] S. Romdhani, V. Blanz, and T. Vetter. Face identification by fitting a 3d morphable model using linear shape and texture error functions. In *Computer Vision – ECCV'02*, volume 4, pages 3–19, Copenhagen, Denmark, 2002.

[12] 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, December 2003.

[13] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. pages 586–591, 1991.

[14] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. In G. Sommer, K. Daniilidis, and J. Pauli, editors, *Proc. 7th Intern. Conf. on Computer Analysis of Images and Patterns, CAIP'97, Kiel*, number 1296, pages 456–463, Heidelberg, 1997. Springer-Verlag.

[15] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, December 2003.