Face and Pose Recognition for Robotic Surveillance

Karl B. J. Axnick
Intelligent Robotics Research Centre (IRRC), ARC Centre for Perceptive and Intelligent Machines in Complex Environments (PIMCE)
Monash University
karl.axnick@eng.monash.edu.au
Ray Jarvis
Intelligent Robotics Research Centre (IRRC), ARC Centre for Perceptive and Intelligent Machines in Complex Environments (PIMCE)
ray.jarvis@eng.monash.edu.au

Abstract

Face recognition has been a huge area of research over the past 25 years [Gao and Leung, 2002]. However the field is still highly unsolved, largely due to variations in pose, illumination and expression. In this paper we propose using pose recognition to solve the first of the face recognition variation problems and simultaneously use both recognised pose and face information to control a robot in a surveillance application. The main advantage of using face and pose recognition to control a robot is a more natural man/machine interaction with the recognised controller (with sufficient clearance) being able to control the robot from their point of view. Also, less manual (hands-on) controls will be required, if the control system is mainly visual. Finally in crowded environments visual control cues from a recognised person are more robust and secure than audible control cues (which are the alternative to either visual or hands-on cues). The recognition method used for both the face and pose recognition is based on geometric 3-Dimensional feature point matching.

1 Introduction

Although pose recognition has already been used to aid face recognition in some instances [Chai *et al.*, 2003], these instances are where 2-Dimensional face recognition is involved. In this paper only 3-Dimensional data will be utilised. By using a 3-dimensional scanner and a texture grabber, a good quality 3-dimensional image can be digitised[1] (Figure 1). This helps solve the earlier listed problems of pose, illumination and expression variation. Illumination variation is overcome in this method as using a laser is an active vision technique that provides independent (invariant) illumination. Expression variation is less detrimental for recognition in this method because 3-Dimensional images have more data points on the face regions which don't vary with expressions compared to 2-Dimensional methods. Finally pose is not a problem in this method as 3-Dimensional face databases inherently have all of the possible poses a target might present, so that once the pose is determined, matching posed images can be generated from the database. Pose recognition will be discussed in Section 2.

Once the face has been localised and the pose recognised, this information can be used to detect salient features on the face. Depending on what features are recoverable and what the normals of those features are, a very accurate pose direction can then be established, beyond the initial pose recognition. This is discussed in Section 3. Although such methods have been tried in previous instances [Elagin *et al.*, 1998], once again those are based on 2-Dimensional data.

Figure 1: A Minolta Vivid 300 scan.

The basic idea in this paper is to enable humans recognised at sufficient clearance levels to control a robot in surveillance applications [Massios and Voorbraak, 1999]. With the anticipation of a future filled with robotic assistances it would be prudent to ensure that the more dangerous of the robotic aids cannot be controlled by criminal or terrorist elements, or simply by someone who

1 http://www.minoltausa.com/vivid/products/vi300-en.asp

does not know how to use such equipment safely. The use of the 3-Dimensional feature points for both face and pose recognition to manage robotic control is explained in Section 4.

The results of many experiments using the methods from the previous sections will be shown in the results Section 5. Finally conclusions will be drawn in Section 6.

Face and pose recognition for robotic control is a novel endeavour. Although pose has been used in previous research [Adachi *et al.,* 2003] to control robots, such instances require very rigid control environments where the controller may as well just use a joystick for robot control. In Adachi the robot is an automated wheel chair for a paraplegic, which will drive the in the controller's gaze direction, if the gaze is held past some time threshold.

The advantage of the method in this paper is that it not only allows the complete mobility of the controller so long as the robot can see their face, but it also ensures that the controller remains consistent and authorised during the control phase. The operator can stand or walk, behind or ahead (facing back) of the robot, to guide it on a new search vector, or the operator could also be within another room with a bank of surveillance screens to guide the robot, if another face scanner was locally mounted in the other room with some remote feed back to the robot. The robot is otherwise an autonomous surveillance machine when it is not being directly controlled by authorised persons, and it follows a path (fixed or random) trying to recognise authorised, unauthorised and criminal persons. If a suspicious event occurs the robot can act on a scripted response or signal for a security guard whilst continuing to monitor the event until a controller enters the scene.

2 Pose Recognition

A hindrance to the current paper's implementation is the requirement for the face under question (being tested for controller clearance) to be between 0.4 – 2.0 metres from the 3-Dimensional scanning equipment, with 0.4 to 0.8 metres being the optimum scanning range for accuracy. A future goal will be to develop 3-Dimensional scanning hardware with a much greater range. Although the current face scans used for generating this paper's results were manually enforced to have the optimum distance, a simple algorithm is proposed in Section 2.1 for automatic scanning of the target only when the target is in the optimal range. Section 2.2 shows how an initial pose direction is inferred using a fast algorithm on the scan results.

2.1 Automatic Scanning

The Minolta Vivid 300 has some intelligence in its operation that must be allowed for. Firstly as already mentioned there is a cube of space that the target must occupy (between the stated range limits to the camera and within the camera's field of view angle) before it can be scanned. Also, the target of the scan must be the nearest object in the scan cube to the scanner, as the scanner assumes the closest object is the intended scan target. To automatically detect that a face is in the virtual scan cube and is the primary object in that scan cube both face detection and range data are required.

Face detection is accomplished at this stage with OpenCV's Haar Face Detector[2] [Lienhart, 2002]. This open source algorithm is very quick and can be trained to find both frontal and profile faces. Figure 2 shows some results of the algorithm. Although pose can be estimated with these results this is not done in this method as such estimations are in the 2-Dimensional image space and are not accurate enough for the current application.

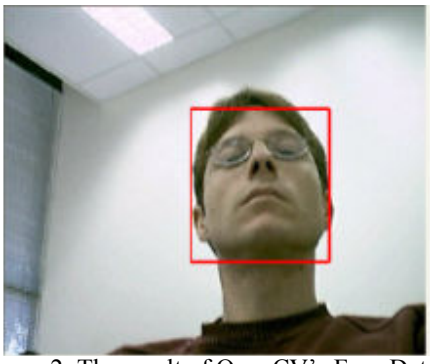


Figure 2: The result of OpenCV's Face Detect

By using two USB cameras to perform the OpenCV face detection, stereo vision can also be simultaneously performed. The method of stereo vision used in this paper is from [Scharstein and Szeliski, 2001][3]. Results of that function are shown in Figure 3.

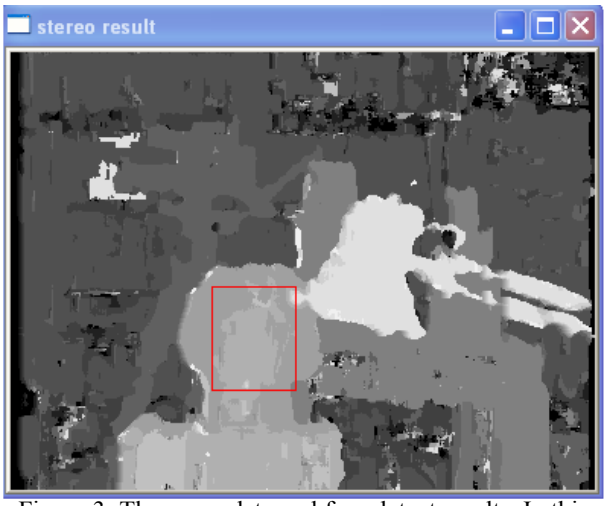


Figure 3: The range data and face detect results. In this case a 3-Dimensional scan was not triggered as the face box is not covering 75% of the closest object, which is a lamp (range image from '3').

If the stereovision result part of the algorithm shows the closest object in the viewing area as being located in (or is overlapping) the region bounded as a face by the OpenCV face detector and the overlap is greater than a certain threshold (75% in current tests), then a 3-Dimensional scan is taken of the detected and ranged face. The only drawback in this algorithm is that the target of the scan is required to be still for 0.4 seconds for noise free results. The result of a clean scan is shown in Figure 1.

2 http://intel.com/research/mrl/research/opencv/

3 http://cat.middlebury.edu/stereo/code.html

2.2 Fast Pose Inferencing

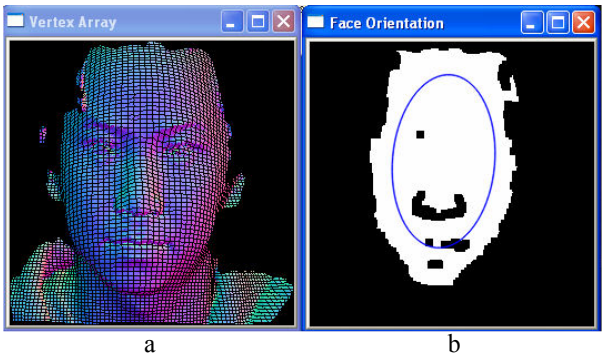
So far progress has been made upon the research efforts of other groups who have been cited and allowed the use of their algorithms for this purpose. This section marks the start of the novel work. Although the scan shown in Figure 1 is very good it can contain well over 100,000 data points. Many of these points are not pertinent for face recognition, such as the hair and shoulders or any other occlusions that may appear in the scan if the person is holding something aloft.

So the first stage in pose recognition after automatic scanning is to automatically cut the extenuous points from the scan and then normalise and filter the useful points left behind. This process is done using only the normals of the 3-Dimensional data points (all colour data is ignored). The normalised normals of points along the x, y and z axis are used to generate a 2-Dimensional colour map of the face as shown in Figures 4a and 5a.

A vertex array is then generated of the points and the average normal difference between neighbouring points is found. A filter then removes all points that have 3 times the average gradient change with any local neighbours. This has the effect of segmenting the 2-Dimensional image into separate point clusters. The remaining points are then dilated so that large blobs are formed where the point clusters were. Then all of these created blobs are labelled and have their areas and moments calculated. The largest blob with the moment that most closely resembles moments of known faces within a threshold is left, while the other blobs are removed (See Figures 4b and 5b). Finally the remaining blob has an ellipse fitted to it. The angle of the maximum axis of this ellipse is then used to predict pose:

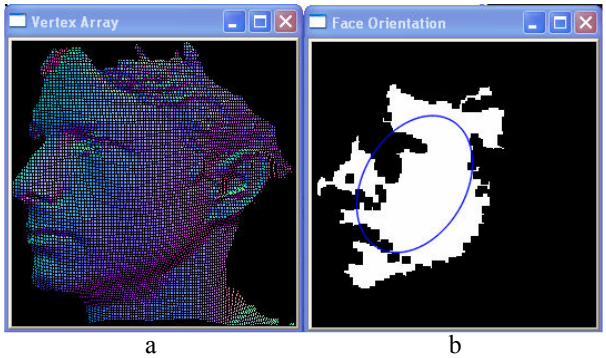
-15º < Angle < 15º → Frontal
Angle <= -15º → Left profile
Angle >= 15º → Right Profile

This pose prediction technique is quite simple and very fast. However, as a forward facing face which was tilted left or right could wrongly be classified as a left or right profile face under the current algorithm; a small complication was added to the above formula, whereby symmetry across the major axis and along the minor axis within the ellipse is measured and a threshold applied to that value to see if the angled face is perhaps forward facing (the threshold returns one if this is the case).



a b

Figure 4: a) Image of normals converted to R,G,B. 4b) Image from Figure 4a after filtering, dilation, blob analysis and ellipse fitting. This one is forward facing.



a b

Figure 5: a) Image of normals converted to R,G,B. 5b) Ellipse indicates that the face is facing right.

3 Feature Extraction

After the face vertices have been isolated and a rough pose estimation has been garnered, a more precise pose recognition can then be realised. By applying many heuristic algorithms the 3-Dimensional surface of the scanned complete/partial face can be quickly traversed to find salient points that are robust to expression variation, impervious to illumination variation and unchanged by pose variation (in terms of relative positions between neighbouring points).

3.1 Normalisation

Before the face surface can be examined however it must be normalised. The first step is to find the average normal direction of all of those vertices remaining. This will give a vector indicating where the scanner was relative to the scanned face. If the pose estimation returned in Section 2 revealed a frontal pose then only a small global rotation to make the scanned face look straight ahead (based on it's average normal direction) needs to be instigated (as the face was already pretty much looking straight ahead). If however as in Figure 5b a right facing pose was found in the Section 2 stage then the normalisation algorithm will need to rotate the scanned face vertices globally clockwise to make the half to three-quartered face (you cannot see the full face surface from a profile scan), face forwards. Monitoring of graphics rendering techniques such as testing that no back culling occurs with the final rotation ensures that the vertices shown at the front of the face mask all those backward-facing vertices at the back of the head. At this point all of the following algorithms can assume a normalised facing face (even if half of the points are missing), see Figure 7.

3.2 Tracing Out Features

One important assumption in this paper is that at the very least, even if only half a face is scanned that this scan contains at least half of a nose (e.g. a quarter is insufficient). The automated scanning process in Section 2 ensures this outcome would be met before scanning is done. This assumption/outcome was also enforced when manual methods were used for generating this paper's data. Given this assumption information it is easy to locate the nose point in the 3-Dimensional mesh, as it is within the point cluster that is closest to the camera. Once this patch is found it is very quickly ratified that adjoining vertices above the supposed nose point slope away slowly and that below the suspected nose point points drop away quickly. This set of local operations ensure that if

somehow jewellery or head ware escapes Section 2.2's filtering that this will not cause incorrect nose locations to be found. If the local filtering finds fault with the suspected nose location then it is rejected and the next closest point cluster to the frontal view is examined.

The local filtering around the suspected nose point in the previous section preludes to how the other salient points are located. Figure 6 shows what points can be searched for and Figure 7 shows what points have been found from the image in Figure 1. The algorithms to find the shown points in Figure 7 are as follows:

1) Starting from the nose point follow the 3D surface along the line of greatest negative change (away from the frontal view direction) until you hit the perpendicular upper lip surface and you have your nose base.
2) Follow the nose ridge along the projection from the nose base to the nose point and when the surface becomes a saddle point you have your nose small.
3) Follow the nose small outwards on both sides seeking wells. The wells are where the eye inners are located on 3-Dimensional faces near the nose small.
4) The Eye outers are found by tracing the valleys above and below the eye line until both paths converge on a well at the other side of the eye.
5) The lips are found by finding the two peaks of the lip ridges below the nose base. The lip ridges can be traced outwards until they merge at the lip outers. There was only one bearded person in the sample set for this paper and the lip finding algorithm was not hindered by his moustache but general moustache robustness for this algorithm is unproven. As the Vivid Laser scanner used tends not to find vertices in hair and as the tracing algorithm for finding lips monitors trends along many points, the few errant points caused by a moustache in a scan should not be confused by the strong clustered vertex concentration on the real lips in the scan.

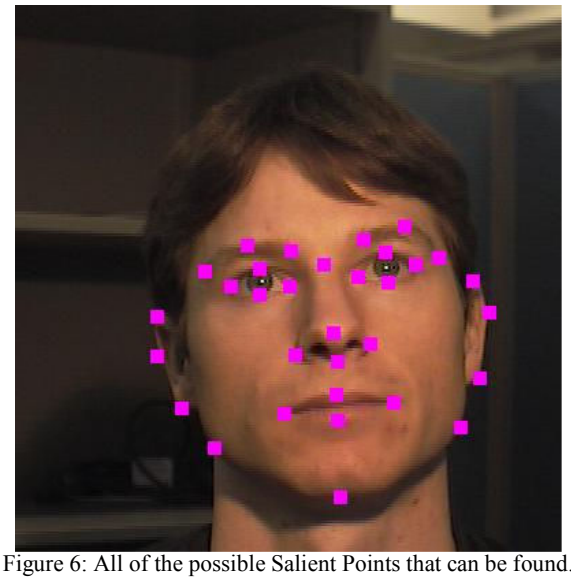


Figure 6: All of the possible Salient Points that can be found.

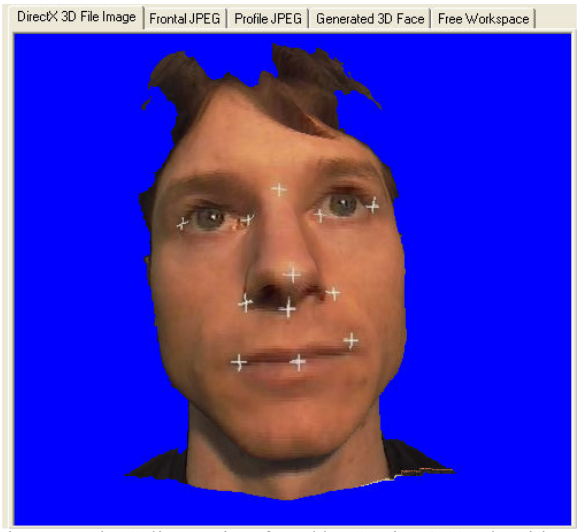


Figure 7: The Salient Points found by Section 3.2's algorithms.

3.3 Pose Recognition

Once as many features that can be extracted from the currently scanned face are found an accurate estimate of the targets pose can be found. Although many other algorithms for pose detection use many point triplet extrapolations of certain features to indicate pose (for example eyes close together indicates a non-frontal view (side on) and the distance from the eye line to the nose gives the looking up or down angle, but as the initial inter-eye width or any other inter-point width is unknown it is unknown how much of the length distortion in 2-Dimensions is caused by the pose, therefore many triplet approximations are needed and only an estimated approximation is achieved) [Heinzmann and Zelinsky, 1998], our paper's method however only needs knowledge of the nose point and the nose base to find the exact pose of the target. It has been found that although all faces are unique [Bruce and Young, 1986] and that what the average normals are for one person's nose may not match any other person's normals for their nose, the vector between the point of the nose base and the nose point is very rigid between all faces, and therefore this measure is the best estimate of which way a person is facing. After this vector is found (using the two nose points mentioned) it is inversely transformed back to where it was facing during the actual scanning, (before the Section 3.1 normalisation). As the robot knows where the scanner is its attention can quickly focus where the current target is looking (using extrapolation) and react if the person being scanned is also recognised as someone who is permitted to control the robot.

4 Face Recognition

Once the exact pose of the face is determined, the identity of the face needs to be recognised (if possible), so that the robot can respond if that person has control clearance, or if that person is a criminal or otherwise unauthorised in the current map location. The method used by this paper for this purpose is Euclidean distance matching between found feature point locations.

As pose variations are rigid body transformations, the relative distances and angles between the salient points are unaffected by pose variations in 3-Dimensions. Also,

as mentioned earlier, since the 3-Dimensional scanning method uses active vision, illumination variations are not bothersome to this paper's methods either. This only leaves expression variation, noise and occlusion as points of failure in this paper's recognition algorithm. This is an advantage not shared by many other competing face recognition algorithms [Hu *et al.*, 2004].

The only remaining sources of error have been minimised in this implementation however to get even better results. Expression variations are combated by giving salient points diminished weights, when they lie in large expression variation areas, most notably the mouth, and other locations learned by differentiating the 3-Dimensonal scans of many test faces with greatly varying expressions but with absolutely no other variants.

Occlusion is removed as a problem in a two fold manner. Firstly, as the method for locating salient feature points scans the local structure of the suspected point locations for prediction confirmation, occlusions such as glasses and hats cannot register as false salient points since hat ridges are not sloped like noses, and the flat surface of glasses certainly have no wells or valleys to traverse, so they cannot be confused as eyes. The feature detection method therefore avoids occlusion variation by ignoring those points that are occluded. The second prong for avoiding occlusion effecting recognition accuracy is in the recognition method itself. If a feature point is not found either because it was occluded or missed through noise, the recognition algorithm basically just ignores that point and such points do not add either positively or detrimentally to the final recognition result. However an addendum to this is that the recognition score from a face with fewer features found has a lower confidence score in its final recognition score, than a recognised face with more feature points. This confidence score doesn't need to be utilised in the current implementation that has a small database of 20 people however. But a larger database might use confidence scores to influence the final recognition score or simply threshold the confidence score before the recognition score is accepted for a target.

The final source of error for this paper's method, which is noise, is handled in much the same way as occlusion was. Whereby, visible points that would have been found but are not found due to corrupted scan data, do not alter the recognition score either positively or negatively, except in the confidence score as previously mentioned. Also the numerous amounts of salient points and the large amount of data known about each point: being normals, absolute position and texture; make the system quite robust to noise that simply modifies the found salient point rather than hiding it.

One final variation that would have been detrimental to this paper's recognition results if not addressed was scale variation. This was handled by the normalisation process of the recognition algorithm which follows:

1) Find the Euclidean point distance between all found salient feature points and crate an array of these values. Substitute a zero for the distance if a point has not been found and also add 1 to the confidence measure.
2) Generate another array which is the result of all Euclidean points' distances from step 1 being divided by each other in order (column 1 is divided by columns 1-32, creating 32 values in the new array). Divisions by zero should be detected and a result of zero substituted for each of the 32 other salient points that would otherwise be divided by the substituted zeros of step 1.
3) Find the Euclidean distance between the normalised Euclidean distance arrays of step 2 with each of the normalised distance arrays stored in memory for all of the known faces that have control privileges. Decrease the distance value involving found points that are vulnerable to expression variation by a learned amount. Increase the found distance value involving points that have been learned to be of high value for recognition accuracy (allow more accurate class classifications).
4) The database face that both has the smallest distance to the current target's normalised distance array, and meets the predicate that the distance is below some threshold is determined to be that target's match and so the robot should proceed to interpret the controllers pose as required. Because the target was in the allowed controllers database. Unless the confidence score of the best match is not below the selected threshold.

The above algorithm is very straight forward and as a result quite fast. The robustness to all of the possible sources of error as explained earlier is the reasons for its design. A final reason for its robustness is that it does not rely on any salient feature point catastrophically. There is no single point of failure, which is present in many other algorithms, such as [Huang and Mariani, 2000], which uses the distance between the eyes to normalise the data, thereby allowing anyone wearing glasses or blinking to break the system. Although not finding the nose point exactly would irk the algorithm in this paper, as long as the nose point location is correct relative to the other salient feature points then the initial error will quickly be reduced by using all of the other points, some of which will be accurate. Also the Section 2 method of automatic good scan verification ensures that at least half of a nose will be always be scanned which is the minimum requirement for successful pose and face recognition. For this algorithm to fail a target would have to have no nose.

Holistic face recognition methods which use entire face images in a grab all manner can easily be corrupted through the earlier mentioned variations [Lai *et al.*, 2001]. Other geometric methods (this paper's method is geometric) such as Gabor Wavelet filters [Sun And Wu, 2002] would however yield more accurate results than this paper's method as wavelet examination of the local area data around salient points is more detailed and robust (as it tests a lot more points) compared to the heuristic methods listed in Section 3. However the increased complexity of such methods is not required for the current application of the face recognition system. Were the database to grow much larger however a less elegant solution than the one currently used would need to be created.

5 Experimental Results

As there are two aims in this research, two separate experiments were required to validate the earlier proposed methods.

5.1 Pose Recognition Results

For the pose recognition experiment, it was considered immaterial whether or not the target was successfully recognised. A wrongly estimated pose for an unrecognised person would not trigger the robot to follow that bad signal, thereby removing the error from the performance of the application. Likewise, a very good pose estimate for an unrecognised person would also not effect the robots performance in the application. In other words, if the face recognition fails, it will not be due to ill-recognised pose. Table 1 shows the results of the pose recognition experiment.

Method	Average Error (%)	Maximum Error (%)
This Paper's	3.7	5.0
View Based Eigenfaces [Srinivasan And Boyer, 2002]	22.1	28.75
Frontal Pose Warping [Jianbo *et al.*, 2000]	28.2	35.0

Table 1: Pose Recognition Experimental Results. For 80 test images from 20 different people.

Each of the people used in the experiment recorded four different poses: left profile, right profile, frontal and random (the angle is known but different people used different angles). These known pose angles were compared with the estimated pose angles for each of the methods listed.

The "View Based Eigenfaces" method basically just uses face detection to recognise pose and is a 2-Dimensional method. By training the Eigenface detector with faces at 0, 15,30,45,60 and 90 degree off frontal poses (in both directions), the pose of the target face can be estimated by the pose value of the trained face detector that showed the strongest result (e.g. if the 15 degree off pose face template finds a face in the image (more strongly than the other templates) then that face must be approximately at a 15 degree off pose angle). Then by using a linear mapping between the value for the strongest responding face pose and the pose values of the faces on either side, a better approximation can be generated. For example if the 15 and 30 degree face templates had similar scores (15's was higher though), and face 0's value was much lower, then the estimated pose would be nearer to 22 degrees, even though the 15 degree face template won.

The "Frontal Pose Warping" method is quite simple, but it is a valid method. It was derived from the listed paper whereby it was understood that although faces differ greatly between people, the feature points undergo similar transformations under pose variations which are rigid body transformations. This means that by mapping 2D corresponding points between two different posed pictures of the same face a transformation matrix to simulate that pose transformation can be generated. By using many different corresponding point pairs an average transformation matrix can be learned for that pose angle. So in this experiment once the nose point is found in the initial 2-Dimensional image grab. By using the nose point as a fulcrum the algorithm rotates the surrounding points around the anchor at different angles (using the different learnt transformation matrices for different pose angles); 0,15,30,45,60 and 90 (in both directions). After each attempted rotation transformation the resultant virtual image is evaluated for symmetry. The rotation which generates the most symmetric face is the rotation that best shows what pose the face was at prior to augmentation.

Once again a linear mapping between angles that have similar scores gives a better final estimate. Obviously a proper 90 degree profile scan will only have half a face that will never be symmetric no matter what the rotation as it is missing its other half, which is why the performance of this method degrades rapidly for poses outside of a 30 degree off centre cone.

This paper's method is significantly superior to the two competing algorithms. This is mainly due to those algorithms being based in 2-Dimensions. The reason no competing 3-Dimensional algorithm was used is simply because there are none of greater accuracy than the simple 3-Dimensional feature point extrapolation of this paper. The 3.7% average error score is based on the assumption of absolute control of the targets' poses during scanning so that their poses would be known. Although angles were measured accurately from the targets centre, with markings on the wall shown to get different people of different sizes and mindsets to all look at the same spot, humans are not perfect and 3.7% would possibly be the natural error of the absolute measurement, making this paper's pose recognition method near perfect.

5.2 Face Recognition Results

The second experiment needed to test the face recognition algorithm. With such a small database of possible allowable people (20) it is relatively easy for the algorithm to find the best match in the database for the current face. However, the trick employed in this experiment was that even though the current face looks most like an approved person, is it actually the approved person? As my database has only approved persons, all targets would always look most like one of the approved persons. To test the threshold used that determines if the best match is actually a valid match the database size was reduced.

By taking 10 peoples' scans out of the 20 person database, these 10 spare scans could then be used to test the threshold for recognition. As the choice of which people were removed from the database could alter the recognition result, it was decided that all combinations of people being removed would be tested and the result was averaged out.

Method	Recognition Rate using only the best match (%)	Recognition Rate using the best 3 matches (%)
This Paper's	100.0	100.0

Eigenface Recognition [Chin and Suter, 2004]	90.3	98.6
LEM [Gao And Leung, 2002]	82.1	88.7

Table 2: Face Recognition results. For 10 people in database and 10 unknown people.

The reason only Eigenfaces and LEM were used as a yardstick for this paper's algorithm, instead of more modern and more accurate methods [Ruiz-del-Solar And Navarrete, 2005], is because the size of the database used in the experiment was small enough, and invariant enough that these lesser but well known recognition methods could give near perfect results. The time and complexity needed to re-implement more modern approaches was unjustified. As speculated earlier, if the recognition results for this paper's method were found lacking it would be easy enough to improve accuracy greatly by utilising Gabor Wavelets on the local area around the salient point locations. Such wavelet results would then allow feature points to be unique in themselves as opposed to only being unique in relation to their relative position to other points. By using the known pose (which is accurately known) to normalise the 3-Dimensional Gabor Wavelet filters convolution around known points it could be possible to recognise a face with only one or a few salient points (as opposed to the 12 currently used points out of a possible 33). This hypothesis will be tested in a future paper that will involve a much larger and standardised database.

The already well known and understood EigenFace recognition and LEM face recognition methods were re-implemented without much change from their respective papers'. However rather than R,G and B being used, the normalised X,Y and Z normals as seen in Figures 4a and 5a were used. This allowed both Eigenface and LEM methods to also be independent of illumination variations, and robust against expression changes; also pose variations could be corrected, to remove them as sources of error. For example if a target had a profile pose, then rather than comparing the generated feature vector with the database frontal pose feature vectors as is, the database images would be rotated to a similar pose to the target and would then have their freshly generated feature vectors compared with the current targets' fresh feature vector. This way no pose variation is present.

The results from this section prove that this paper's method is superior to well entrenched simple face recognition methods. As our method scored perfectly it is just that one could hypothesise that such perfection would continue as the database grew, until the variation between the 33 possible salient feature points in 3-Dimensions (with 9 parameters at each point, X,Y, Z, normals and colour) could not uniquely encode every face in the database. However when this threshold is breached, it has been noted that using a Gabor Wavelet Filter (or any filer in fact) on the local area around points will give virtually endless uniquely coded parametric results. At the same time methods that do not give perfect results on this small database will continue to degrade as database sizes increase, making Gao's and Chin's methods obsolete.

6 Discussion and Conclusions

The purpose of the paper was to develop algorithms that perform face and pose recognition in a security function, whilst simultaneously using pose and face recognition to allow an operator to control a robot remotely, accurately and securely in that surveillance application as needed. Although for the most part the robot is automatic in operation, simply randomly roaming or patrolling set paths and scanning people where it can to find both unauthorised persons and security persons, there are instances when control will be needed by human operators. In such cases the robot could be programmed to pause its current function when an authorised person looks straight at it, and then take further action based on what the operator then conveys with further poses or manual input. If an authorised person does not look at the robot then a robot will continue its automatic operation around the authorised person.

Although pose recognition is not required for this paper's 3-Dimensional face recognition algorithm, it is an inherent by-product of the method and should be utilised if productive to do so. Even though an authorised person could always stop a robot under its control either with a network signal or a secret code/button on the robot, these methods are easily fallible by any skilled antagonist. Also there is a certain ineloquence with manual manhandling of systems/robots that when avoided makes for a more professional application. A simple yet highly useful extension of the pose and face recognition methods demonstrated in this paper is at the ATM. If a customer not only needs a PIN but also the correct face, and a subtle pose when they enter the PIN, the system will be vastly more secure. An armed kidnapper forcing a person to enter details would be none the wiser if a person looked slightly left when entering bank details under duress to make the ATM seem cooperative to the robber but show a balance of zero in the account as per the customer's subtle command. With remote banking growing exponentially any extra anti-fraud mechanism is welcome.

The results from Section 5 proved that our system's pose recognition is extremely accurate and so this part of the paper's aim was readily met. However the relationship between a clear signal (a correctly classified pose angle) and a robots response (to the recognised pose) is valued differently in different applications. For instance if a robots range of motions are only forward, left or right then an accurate distinction between a 60 degree off centre pose to the left and a 70 degree off centre pose to the left will not improve the performance of telling the robot to turn left.

However, as humans use body language to convey 80% of communication it is essential that robots evolve to a level where even subtle body angle adjustments can be translated into reactionable algorithms by robots/machines in the near future. This research paper is hopefully a first step towards this goal. The pose recognition accuracy although redundant in many instances will allow for creative use, as in the ATM example above.

Even though this paper has focussed on the more elaborate, yet subtle creative uses for pose recognition based control, such as covert control messages (a novel

area); the more obvious and overt simple direct control interpretations are just as apt for pose recognition control methods. For instance: using the up/down nod of the head for forward/backward robot movement; using turning the head left/right to make the robot turn left/right; and finally using tilting the head left/right to make the robot pan an independent (of the robot's main body) directional camera or manipulator left/right. However, perhaps the ultimate benefit of head pose based control would be merging it with manual methods. The operator could use gaze direction to rapidly centre a robot's manipulator (globally *e.g.* the wrist and shoulder) at certain locations, and then use a manual controlling device (*e.g.* a joystick) for fine manipulations (locally *e.g.* the hand and fingers) centred on the gaze target point (which the controller would need to look at anyway), to manipulate end effectors such as a grasping device. There should be no limit to the creative uses of pose recognition based controls.

The face recognition accuracy result from the experiment in Section 5.2, like the pose recognition section was also very good as it proves that the last aim of the paper was achieved. However, as face recognition and face validation are quite distinct in complexity, it should be pointed out that the current implementation for this paper is more of a face validation scenario than a face recognition scenario, due to the size of the database.

Future work is being conducted whereby the database is much larger, but still with a small subset of authorised users. This will allow the robot to be more than just a receptive tool. Such a robot could automatically patrol locations looking not just for controllers but for criminals or black listed individuals in its database, a more face recognition orientated class. When such a person is found the robot could call human security to help with the situation, then the ability to discern if the security guard is a valid controller will come to the fore. In such a confrontation the ability of the robot to read subtle pose changes could be used to convey hidden meaning that even a fellow human security guard could not read. For instance, if voice recognition is also used as a robotic control, a pose of 10 degrees off pose to the left could mean do the inverse of what the controller verbally commands. So if the security guard wished the suspect followed without giving this away he could verbally command with the negating pose saying "This person is cleared, do not follow them further". Of course the nonverbal subtleties do not need verbal confirmations. A simple 10 degree off centre pose to the left for 10 seconds, could issue the same command without any verbal cue.

So although face recognition is the main goal of the paper and it was achieved, the novel spin off of such an accurate algorithm is the effective use of the pose recognition by-product. As demonstrated in Section 5.2 although pose variation does not hinder the current face recognition method, if the method was expanded to use for example Gabor Wavelet filters around the salient points, then pose knowledge is critical for normalisation. Also knowing the pose allows other proven face recognition methods that are vulnerable to pose variations to be used in this application, as the 3-Dimensional database and pose recognition allows pose variations to be removed (as seen in Section 5.2).

References

[Adachi *et al.*, 2003] Estimation of user's attention based on gaze and environment measurements for robotic wheelchair Adachi, Y.; Goto, K.; Khiat, A.; Matsumoto, Y.; Ogasawara, T.; *Robot and Human Interactive Communication*, 2003. *Proceedings.* ROMAN 2003. *The 12th IEEE International Workshop on* 31 Oct.-2 Nov. 2003 Page(s):97 – 102

[Bruce and Young, 1986] Bruce, V. & Young, A. (1986) Understanding face recognition. *The British Journal of Psychology*, 77 (3), 305-327.

[Chai *et al.*, 2003] Pose normalization for robust face recognition based on statistical affine transformation Xiujuan Chai; Shiguang Shan; Wen Gao; *Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the* 2003 *Joint Conference of the Fourth International Conference on* Volume 3, 15-18 Dec. 2003 Page(s):1413 - 1417 vol.3

[Chin and Suter, 2004] T.J. Chin and D. Suter, MECSE-6-2004: A Study of the Eigenface Approach for Face Recognition. June 2004, *IRRC, ECSE, Monash University*.

[Elagin *et al.*, 1998] Automatic pose estimation system for human faces based on bunch graph matching technology. Elagin, E.; Steffens, J.; Neven, H.; *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on* 14-16 April 1998 Page(s):136 – 141

[Gao and Leung, 2002] Face recognition using line edge map, Yongsheng Gao; Leung, M.K.H. *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, Vol.24, Iss.6, Jun 2002 Pages:764-779

[Heinzmann and Zelinsky, 1998] 3-D facial pose and gaze point estimation using a robust real-time tracking paradigm Heinzmann, J.; Zelinsky, A.; *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on* 14-16 April 1998 Page(s):142 – 147

[Hu *et al.*, 2004] Hu, Y., et al. Automatic 3D reconstruction for face recognition. in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on.* 2004.

[Huang and Mariani, 2000] Face detection and precise eyes location Weimin Huang; Mariani, R.; *Pattern Recognition, 2000. Proceedings. 15th International Conference on* Volume 4, 3-7 Sept. 2000 Page(s):722 - 727 vol.4

[Jianbo *et al.*, 2000] Recovering frontal-pose image from a single profile image Jianbo Ma; Ahuja, N.; Neti, C.; Senior, A.W.; *Image Processing, 2000. Proceedings. 2000 International Conference on* Volume 2, 10-13 Sept. 2000 Page(s):243 - 246 vol.2

[Lai *et al.*, 2001] J. Lai, P. Yuen, G. Feng, "Face Recognition Using Holistic Fourier Invariant Features", *Pattern Recognition*, 34(1), 2001, pp95-109.

[Lienhart, 2002] Rainer Lienhart and Jochen Maydt. An extended Set of Haar-like Features for Rapid Object Detection, *IEEE ICIP 2002*, Vol 1, pp. 900-903 Sep. 2002.

[Massios and Voorbraak, 1999] Hierarchical decision-theoretic planning for autonomous robotic surveillance Massios, N.; Voorbraak, F.; *Advanced Mobile Robots, 1999. (Eurobot '99) 1999 Third European Workshop on* 6-8 Sept. 1999 Page(s):219 – 226

[Ruiz-del-Solar and Navarrete, 2005] Eigenspace-based face recognition: a comparative study of different approaches Ruiz-del-Solar, J.; Navarrete, P.; *Systems, Man and Cybernetics, Part C, IEEE Transactions on* Volume 35, Issue 3, Aug. 2005 Page(s):315 - 325

[Scharstein and Szeliski, 2001] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Technical Report MSR-TR-2001-81, Microsoft Research*, November 2001. which can be found at http://www.research.microsoft.com/scripts/pubs/view.asp

[Srinivasan and Boyer, 2002] Head pose estimation using view based eigenspaces Srinivasan, S.; Boyer, K.L.; *Pattern Recognition, 2002. Proceedings. 16th International Conference on* Volume 4, 2002 Page(s):302 - 305 vol.4

[Sun and Wu, 2002] A local-to-holistic face recognition approach using elastic graph matching Da-Rui Sun; Le-Nan Wu; *Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on* Volume 1, 4-5 Nov. 2002 Page(s):240 - 242 vol.1