# On The Synergies Between Online Social Networking, Face Recognition And Interactive Robotics

Nikolaos Mavridis and Wajahat Kazmi
Interactive Robots and Media Lab
United Arab Emirates University, 17551
Al Ain, United Arab Emirates
E-mail: irmluaeu@gmail.com

Panos Toulis
ISSEL laboratory, Aristotle University of Thessaloniki, Greece
Email: ptoulis@olympus.ee.auth.gr

Chiraz Ben-Abdelkader
New York Institute of Techology
Abu Dhabi campus
United Arab Emirates
Email: chiraz@nyit.edu

***Abstract*—This paper explores the intersection of three areas: interactive robots, face recognition, and online social networks, by presenting and discussing an implemented real-world system that combines all three, a "FaceBots" robot. Our robot is a mobile robot with face recognition, natural language dialogue, as well as mapping capabilities. The robot is also equipped with a social database containing information about the people it interacts with and is also connected in real-time to the "Facebook" online social networking website, which contains information as well as partially tagged pictures. Our system demonstrates the benefits of this triangle of interconnection: it is not only the case that Facebook information can lead to more interesting interactions, but also that: Facebook photos enable better face recognition, interactive robots enable robot-mediated publishing of photos and information on Facebook. Most importantly, as we shall see in detail, social information enables significantly better and faster face recognition, as an interesting bidirectional relationship exists between the "friends" relation in social networks and the "faces appear in the same picture" relation in face recognition. We will present algorithms for exploiting this relationship, as well as quantitative results. The two main novelties of our system are: this is the first interactive conversational mobile robot that utilizes and publishes social information in Facebook, and is also the first system utilizing the social context of conjectured identities in a photo for better face recognition.**

## I. Introduction

THE main problem originally intended to be addressed by the "Facebots" project ([1],[2]) was that of the creation of sustainable and meaningful long-term human-robot relationships. This is a most important problem towards our ultimate goal of human-robot symbiosis, i.e. harmonious and mutually beneficial living together of the two species. And also, in the shorter term, this is an important problem towards the successful application of robots to numerous areas: disabled and elderly assistance / companionship, supporting education, and more. Furthermore, current results illustrate that we have not advanced significantly towards sustainable long-term human-robot relationships (for example, [3]). Our proposed solution to this problem, was "Facebots", robots that can create better long-term relationships with humans, through reference to past meetings with them ("shared memories"), as well as to their friends ("shared social circle"). The key towards implementing such systems was the creation of interactive robots equipped with an online interaction database as well as social database, which furthermore was connected online to the highly-successful social networking website "FaceBook". "Facebots" robots are described in [2].

Here, after discussing background literature and providing a very brief overview of "Facebots", we will move towards a tangential, yet more general and quite important direction, which came up as a strong theme while developing "FaceBots". Three main areas were implicated from the beginning in the creation of our robots: interactive robotics (IR) face recognition (FR), as well as online social networking (SN). However, our original and main problem to be addressed was that of the creation of sustainable and meaningful human-robot relationships; i.e. we were using face recognition as well as online social networking towards better interactive robotics. But soon it started to become clear that this (i.e. FR and SN aiding towards IR) was not the only direction and combination of benefits in our triad of main areas.

In this paper, we will discuss a number of other such directions and combinations providing benefit. As a first combination, we will illustrate how interactive robotics and social networks can aid face recognition (IR and SN aiding FR). For example, this can be achieved by utilizing social information (friendship relationships) towards beneficially biasing recognition in pictures, and achieving higher recognition accuracies and decreased recognition time. The underlying idea enabling this is use of social context: the relation "A and B appear in the same photo" is highly correlated with "A and B are friends". In another example that will be discussed, the benefits of online social networking to face recognition are also illustrated by utilizing Facebook-derived tagged pictures to augment small training sets of robot-camera-derived pictures in order to enhance recognition. As a second combination (and as a third example), we will discuss how interactive robots and face recognition can aid online social networking. We hope that this paper will serve as an illustration of the strong synergies between the three areas, and will aid towards their further co-development.

Manuscript received April 30, 2009. This work was supported in part by the Microsoft USA External Research Department, under its social robotics call for proposals.

## II. BACKGROUND

### A. Related Work

Numerous examples of interactive robots (IR) exist; for example (Kismet [4], Leonardo [5], Maggie [6], Robovies [7] and more). However, no existing systems have utilized online social networks. Thus, Facebots are the first system that uses (SN) towards (IR). Also, social information (knowledge of the web of friendships etc.), has not been utilized towards automated face recognition yet, and thus what we will describe is the first system that uses the social information component of (SN) towards (FR).

Nevertheless, using face detection and face recognition (FR) towards interactive robots (IR) is not a new idea; there are numerous projects built-around face-detecting robots [8],[9], which might even carry out conversations with multiple humans, such as in [10].

Few attempts exists towards utilizing face recognition (FR) on images belonging to online social networking (SN) websites (for example [11], without utilization of context). On the other hand, methods relevant to context-assisted face recognition have also appeared in the literature recently: [12] provides an example of contextual priming for object recognition, based on holistic context representations, while [13] performs object detection by modeling the interdependence of objects, surface orientations, and camera viewpoint. However, none of these papers address the utilization of social context for face recognition; the only noteworthy exception is [14]. There is an important difference though between this paper and what we are presenting here: [14] only uses the identity of the person contributing the photo to the online networking website in order to enhance the recognition, and the method only works if this is known. In contrast, our method does not require this information; it can be seeded by the social context created through postulated or recognized participants in the photo, and is much more flexible in that respect, and can thus be used also on photos with no submitting author information, arising anywhere on the internet or live.

Regarding the creation of sustainable human-robot relationships (IR), a key long-term (six month) study is [3]. Shorter field studies in other contexts have taken place in the past; for example the 18-day field trial of conversational robots in a Japanese elementary school [15]; and numerous are underway, including a possible massive deployment of humanoids in malls [Ishiguro, personal communication].

Finally, apart from utilizing online social networks toward interactive robots, one could envision the real-time utilization of other web resources by such robots. Exciting such prospects exist; for example, “Peekaboom” [16], could in the future serve as a real-time repository for object recognition, situationally-appropriate interactions could be learnt through experience arising from online virtual worlds, and much more.

### B. The Facebots Robots

For more details, refer to [2]. An overview follows:

***Hardware:*** Our robot is composed of an ActivMedia PeopleBot robot [17], augmented with a SICK laser range finder, a touch screen, and a stereo Bumblebee camera [18] on a pan-tilt base [19] that is at eye-level with humans. For a picture, look at figure 1.

***Software:*** We have created an expandable modular software architecture, with modules intercommunicating through the ICE IPC system [20]. The modules can be running on multiple CPUs or PCs which are part of a network, and are written in C++, Java, and Perl. Effectively, a callable-method API is exposed by each module towards the others. The modules we have created are:

(M1) *Vision Module* with Face Detection & Recognition, from camera- or Facebook- derived pictures. Includes real-time externally callable training set modification / new classifier generation capabilities, and pluggable face detectors / classifiers. Uses an implementation of [21] for face detection in conjunction with a skin-color based method, [22] for pre-processing, and a variation of [23] for recognition.

(M2) *Natural Language Dialogue Module*, with real-time language model switching capabilities. Uses [24] for recognition, and the *Cepstral* module of [25] for synthesis.

(M3) *Social Database Module*, which locally holds basic personal info / friendship relationship / simple event data / photos for the people the robot knows, and which connects and updates through Facebook for those that are members of it. This module also contains an interaction database.

(M4) *Navigation and Motion Module*, that can build a map of the robot’s environment, drive to key social locations and drive the camera’s pan-tilt unit

(M5) *Controller Module*, which issues calls to all other modules, and where high-level system operation routines can easily be scripted.

Fig. 1. Our robot, Sarah the Facebot, while interacting with a human

***Operation:*** The robot wanders in its environment until it detects a human face in proximity. Then, recognition is attempted, and if there is enough confidence, the human is greeted and his identity is verified. A typical session follows:

*//Greetings / Recognition:*
H:Hello! My name is Sarah. Welcome to the interactive robots and media laboratory
R:Hi! Are you Panos Toulis?
H:Yes (Robot changes its online status to "interacting with Panos Toulis")
*//Query of overall state:*
R: Hey, Panos Toulis, are you doing well today?
H: Yes, pretty well
*//General news-based discussion:*
R: Let me tell you some interesting news that I've heard
R: The latest windows release will be available to everyone after a surge in demand crashed the microsoft website on the 9th of January (item taken from BBC news online)
R: Have you heard about this?
H: Yes
R: That's great!
*//Robot discusses Facebook status of human*
R: I see you are chasing robots in the desert
*//Robot gives news about a mutual friend, acquired through Facebook*
R: Did you know that our friend Shervin Emami has posted a new photo on Facebook? (information acquired from Facebook minifeed)
H: No
*//Robot sends reminder to human*
R: I am sending you a message about this to check it out Robot remembers past encounter with mutual friend
R: I saw Shervin yesterday evening. He was doing well. (information acquired from interaction database)
*//Robot gives news about a mutual friend, acquired through Facebook*
H: Did you know that our friend Chandan is "waiting for Sarah to have that little sweet conversation with me"? (information acquired through chandan's Facebook status)
R: Yes, of course.
*//Robot connects an online friend with the human*
R: One of our friends, Nikolaos Mavridis, is online – would you like me to send him a message?
H: Yes ('Nikolaos, Panos says hello" is sent)
*//End of interaction*
R: Hey, Panos, it was nice talking to you!
R: I have to go now.
R: See you later!

***Tuning and Evaluation:*** A number of basic tuning and evaluation experiments are discussed in [1]. These include the tuning of thresholds for classifying a face as "unknown", of the number of frames for temporal evidence accumulation for face recognition, of the training time required as a function of training set size, as well as some evaluation results from an early human-robot interaction experiment.

Most importantly, the question of transferability of training from camera pictures to Facebook photos and vice-versa is touched upon, and some initial quantitative results are given. Here, we will elaborate on this question, extend our results, and comment upon them in our context of synergy within the interactive robots – social networks – face recognition triangle.

Finally, the possibility of utilizing social information (friendship networks) towards face recognition was very briefly introduced in [2]. A detailed quantitative and mathematical exposition and description of the method developed is presented here, within the context of the synergies of our triangle.

## III. MUTUAL BENEFITS

Three examples of mutual benefits are presented here in detail. First, social information assisting face recognition. Then, online social networking site-derived pictures complementing robot camera-derived and vice-versa for face recognition. And finally, interactive robots providing information for online social networking sites.

### *A. From Social Information to better Face Recognition*

Here we will discuss how one can utilize Facebook- or robot-derived friendship information, in order to enhance automated tagging in Facebook pictures. The underlying assumption is that friends are more likely to co-occur in photos (and we will model and quantify this below). Thus we can start biasing our recognition hypothesis set towards friends, once we know the identity of a person in a photo. The proposed process goes as follows:

Suppose we know the identity of a person, either through recognition, or through pre-tagging, and that we are quite confident of it. Then, we acquire his circle of friends through the social database, and we bias our hypothesis space (bigger priors, larger score weight etc.) towards the circle of friends. After that, we perform recognition of the other faces, and choose the one whose identity we are most confident of.

Now, we have two circles of friends: the first face's friends (F1), and the second (F2). We also have their intersection: their mutual friends (F1&2). Thus, we can now bias with three levels of strength: one for non-friends (not belonging to either F1 or F2), another for mutual friends (F1&2), and yet another for friends of F1 or friends of F2 which are not mutual. This is the overall idea. Let us now move on to a more detailed description and specifics.

***Relevant statistics:*** The given name of the first "Facebot" that we have built is "Sarah Mobileiro", which is also her online name. The robot so far has 79 Facebook friends, out of which 14 she has met physically, and has also acquired camera pictures of. The robot also has another 80 friends who are not on Facebook, and also has camera picture of. The set of the 79 first level friends (direct friends) of the robot in Facebook, we will from now on call FL1 (see fig 2).

Upon moving from the first level friends to the second level, i.e. the friends of the first level friends of the robot who are not first level friends, there is, as expected, a huge increase: the set FL2 (of friends with minimum distance 2) of Sarah the Facebot contains 13989 members. By a simple division, one gets the figure of on average approximately 177 new second level friends for each first level friend. Of

course, the average number of second level friends corresponding to each first level friend is higher (211 as compared to 177, i.e. on average 34 friends are shared, i.e. approximately 15% of the friends are shared). This is due to the existence of mutual second level friends between any two first level friends. Also, it is worth noting that the variance of the number of friends of each member of FL1 is quite high too – 122 in this case.

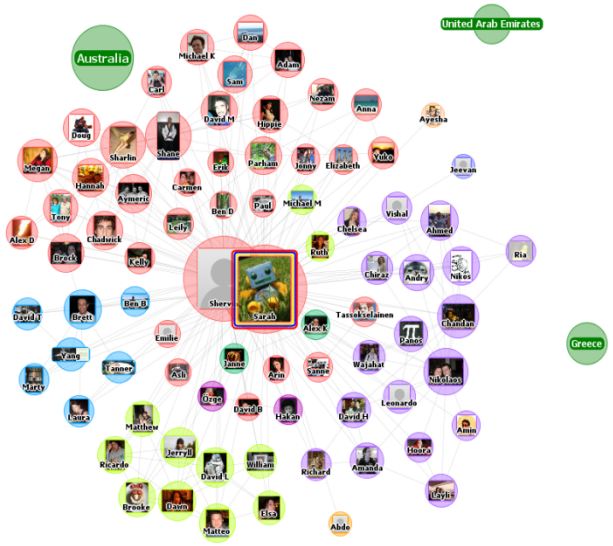


Fig. 2. The "touchgraph" depiction of the first level friends of our robot

All the above statistics are related to the social network of Sarah, at maximum distance two. Now, having briefly explored this, let us move on to the next question in sequence: how many of the first- and second- level friends of Sarah can we create classifiers for the face recognition of?

The total number of tagged photos of the members of FL1 and FL2 which are directly accessible to Sarah is 11209. This number arises as the sum of the number of tagged photos across each first-level friends – tagged photos of second-level friends are not generally accessible due to visibility constraints). The distribution of the number of available tagged photos for the 79 first-level friends is given below in figure 3.

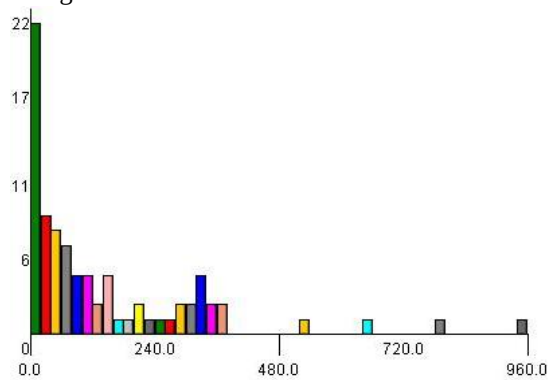


Fig. 3. Histogram of number of available tagged photos per first-level friend of the robot (these tagged to be potentially utilized as a training set)

The average number of tagged photos per first level friend is 141, with a standard deviation of 179 – easily explicable through the 4 outliers with more than 300 tagged photos (see Figure 1). Thus, we expect to have a wide variety of training set sizes – numerous friends have only 1 photo available, while a significant number might have 100 or more.

Now, although as we mentioned there is a sum of 11209 photos when tagged photos are summed across the 79 first-level friends, not all of these are unique. Out of these, given the possibility of a photo having more than one person tagged, the number of unique photos is 7647, including 44 problematic images, leaving 7603 usable. Furthermore some of these will have only a single face tagged – and some more than one face. Indeed, more than two thirds out of the 7647 unique photos have more than one tagged face, as can be seen from the histogram of number of photos containing n tagged faces in figure 4. And now the question arises: for how many of the first- and second-level friends of the robot do we have adequate training sets to create classifiers out of? If we restrict ourselves to gathering training data through these tagged photos (the simplest and safest solution), then we have at least one tagged photo for only 3595 out of the 79+13989 = 14068 first and second level friends of Sarah, i.e. roughly 25% of FL1 and FL2.

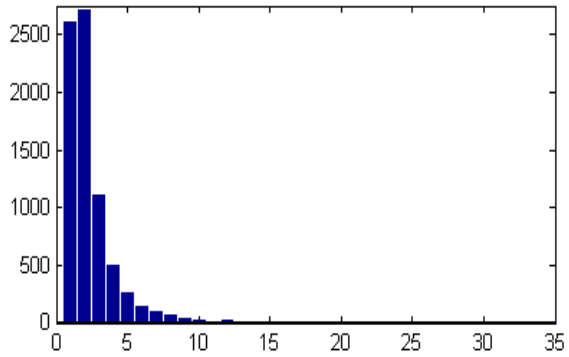


Fig. 4. Histogram of number of photos containing exactly n tagged faces

***Two problems and two solutions:*** So far we have superficially examined the first- and second- level social network of Sarah, as well as the availability and sizes of training sets for recognizing the faces of her friends. Now, we need to move on closer to our goal: the utilization of social information towards better face recognition. Let us do so by introducing a *concrete problem:*

(Pro1): Assume that we have a partially tagged photo with n detected faces in total, in which the identity of only one face is known, while there are (n-1) faces of unknown identity. Our goal is to automatically tag the n-1 remaining unknown faces.

A *simplistic solution* to the above problem, which would not utilize social information, would be the following:

(Sol1): Take each unknown face in turn, and apply all of our classifiers to it. Examine the vector of confidences for each possible identity, find the maximum confidence and the corresponding most likely identity, and if this maximum confidence is above a minimum acceptable threshold, tag the face with this most likely identity. Else, tag the face as "unknown".

In the case of Sarah, this would require going through

3595 classifiers for each unknown face, each of them considered as a priori equally likely (as a first approx.).

Now, let us try to improve on this, by incorporating knowledge of social information. Let us consider the following *three questions*, and try to obtain empirical answers for them:

Q1) Given a person A in a photo, what is the probability of any other person B in the photo being a first level friend of A? Let us call this P1.

Q2) Given a person A in a photo, what is the probability of any other person B in the photo being a second level friend of A? Let us call this P2.

Q3) Given two persons A and B in a photo, what is the probability of any other person C in the photo being a mutual friend of A and B (where it is not necessary that A is a first level friend of B)? Let us call this P3.

By examining all the 5214 unique tagged photos with more than one tagged face that Sarah has direct access to, we obtain the following estimates for the three above probabilities (measured across tagged photos with >1 face):

P1 ~ 0.785 P2 ~ 0.024 P3 ~ 0.278

P1 is strikingly high: almost 80% of any two faces in photos are first-level friends. Now, let us try to incorporate the above empirical estimates in our classification scheme, by biasing our classifiers. Consider the following algorithmic solution to Pro1:

(Sol2) *Step1*: Let us call the known face F1. Take the circle of first level friends of the known face (F1FL1), as well as the circle of the second level friends of the known face (F1FL2). Now, consider an unknown face and apply all available classifiers, but bias their confidence output through a multiplicative weight (additive in the case of log prob). This weight will carry a higher value for F1FL1, an intermediate value for F1FL2, and a smaller for other classifiers. Examine the vector of confidences for each possible identity, find the maximum confidence and the corresponding most likely identity, and if this maximum confidence is above a minimum acceptable threshold, tag the face with this most likely identity. Else, tag the face as "temporarily unknown". If the face was marked as "known", then keep track of its confidence, and move to Step2. Else, repeat Step1 with another unknown face chosen.

*Step2:* Now, two faces are currently known: one which was initially tagged as known (F1), and one that we have autotagged so far (F2), through step1. Thus, we have the following circles of friends which can be derived through our known social information: first level friends of F1 (F1FL1), second level friends of F1 (F1FL2), as well as similar sets for F2 (F2FL1 and F2FL2) and mutual friends of F1 and F2, which belong to the set which we will call F12MF. A more detailed analysis will uncover a partitioning of 3 x 3 x 3 = 27 mutually exclusive possibilities (3 arising out of {first level, second level, none}, Cartesian product of which is taken twice to account for F1, F2, and the new unknown face X). A closer examination shows that 3 out of the 27 combinations are impossible (for example, {F1 first level to X} while {F2 first level to X} while {F1 not related to F2} is impossible, because it follows by the two premises that it should be the case that {F1 is at least second level to F2}). Now, the set of classifiers of all faces is partitioned to these 27-3 = 24 categories, appropriate weights are given to each category (highest for mutual friends, high for first level of one only, medium for second level of both, lower for second level of one only, lowest else), and if the winning classifier has an acceptable confidence, we tag it as "known" and keep track of its confidence. Else, we tag it as "temporarily unknown" and return to Step2. If we have tagged the face as "known", a third step which is a generalization of the Step 3 for three known faces follows, and so on.

Multiple variations of the above scheme are possible: first, there is the possibility of backtracking in case a particular sequence of decisions for "temporarily assumed known" faces does not satisfy a metric of overall high confidence as well as minimum confidence bound. Also, in order to reduce total computation time in the case of a large number of classifiers, one can temporarily totally discard the search through the classifiers which have lowest weight; and only keep those belonging to one of the friend's circles. Furthermore, to avoid computational complexity, one can proceed with multiple runs of step 2 without ever moving to step 3 or further, by choosing the two faces with highest confidence as seeds.

Thus, the algorithm presented in (Sol2), will enhance face recognition, given the proper choice of weights, which are related to the probabilities P1, P2 and P3 estimated above. Depending on the type of confidence measure given by the classifiers, these weights can be chosen either empirically or analytically. Furthermore, the above algorithm Sol2 can also decrease computational cost, if we use the discarding option discussed above.

***Specifics – Algorithms and methods***

On the basis of the previous discussion, the following specific experiments were carried out, to investigate the effectiveness of incorporation of social information in the face recognition process. There are three main species of objects involved here: *photos* (including multiple faces), *faces* (i.e. cropped regions containing a face), and *persons* (tags/identities/friends). Below, we will see how the photos were split to photos comprising a training set and photos comprising a testing set, how we performed face detection and extraction, how classifiers were created, and how testing took place using a variety of algorithms, some of which were also utilizing social information.

Out of the 79 first level friends of Sarah, 9 had no tagged photos. As mentioned before, these contribute 7603 usable unique tagged photos. These were split to two sets randomly. Then, *face detection*, matching to tags, and extraction was performed to these, in the following manner, using the Viola-Jones detector [21]:

1-The face detector returns face regions in the form of

rectangles with top left corner coordinates and height and width given in pixels. The center of the face region is then calculated.

2-The Facebook face tags are accompanied only with a center-point expressed in percentage of image height and width. This is transformed to center point coordinates.

3-A face detector regions – to – Facebook tags matching procedure takes place, based on horizontal and 2D distances of centers.

4-Preprocessing (histogram equalization, resizing) takes place and the normalized crop face is saved.

The *training set* contained only 2809 photos containing detected faces (our face detector only works for frontal faces and after correct matching of tags). In these photos, 1306 unique tags appeared (including people without Facebook profile), in a total of 5157 cropped face images (3.94 faces per person average). These were used as our training set for 1306 classifiers (embedded HMM-based as in [23]). However, although there were nearly 4 faces per person on average, the individual distribution of number of faces available for training every person reveals that more than a half of the classifiers had a single photo as a training set, and a quarter two photos; only the remaining quarter had more than two (see figure 4). Training took around 2.5 hours on a Xeon Processor (Quad Core, 2.66 Ghz each core), with 3GBs of RAM. Every face image together with its mirror image was used for training.

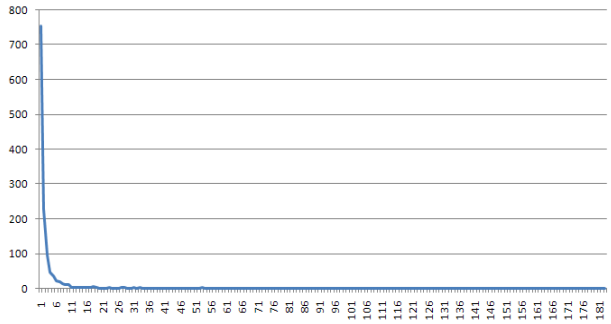


Fig. 4. Number of photos containing exactly *n* tagged faces

The *testing set* contained only 2765 photos containing detected faces (our face detector only works for frontal faces and after correct matching of tags). In these photos, 1372 unique tags appeared (including people without Facebook profile) in a total of 5258 cropped face images (3.83 faces per person average). These face images were used as our testing set. Testing one face against 1306 classifiers took roughly 28sec on our Xeon machines, thus the total estimated testing time if running on a single CPU was of the order of 2 days. Therefore, a server was set up, distributing face images for testing to multiple machines, 3 dual core (2.13GHz per core) and 4 Xeon processors (2.66GHz per core), giving an effective 22 CPUs, reducing testing time to a little more than 2 hours.

Three algorithms were used for quantitative testing: the first one (Alg1) assumed one face per photo known (problem pro1), but was not utilizing social information at all (providing a baseline for comparison). The second (Alg2) assumed one face per photo known (again pro1), but was utilizing social information. The third (Alg3), assumed no faces known at all initially, and utilized social information in the recognition process.

A very important point not discussed yet deals with the *amount of overlap* between the identities (people) included in the training set and having formed classifiers, and those tagged in the testing set. As mentioned above, there were 1306 classifiers and 1372 unique tags in the testing photos; however, only 379 people had classifiers and appeared in the testing photos – i.e. only roughly a quarter of the people appearing in the testing set we had classifiers for.

The demographics of the identities (people) are also quite interesting in their own right. As mentioned above, the *intersection* of training and testing set identities is 379, and the *union* of the training and testing set identities is straightforward to calculate: 1306+1372-379=2299 people. These can be divided into *five categories*: those belonging to the first level friends of the robot (F1), those belonging to the second level friends of the robot (F2), those who are on Facebook but not first- or second-level friends (F4), and those who are not on Facebook (F5). Rough percentages of these categories within the union and intersection follow:

Union: F1 3%, F2 55%, F3 28%, F4 14%

Intersection: F1 16%, F2 69%, F3 10%, F4 5%

Finally, it is worth noting another point, before we report the recognition results of our novel method which incorporates social information. In the previous paragraph we mentioned that there is *partial knowledge of the extended friends network*; this is the case due to the partial visibility setting of Facebook, that users can set. Our algorithm for exploring friends that was used here in reporting the above percentages was the following: We take each photo belonging to the training and testing sets in turn, and consider each tagged identity in them (i.e. the people belonging to the Union of identities mentioned above). For each one of them, we try to get their friends list if it is open. Then, we concatenate all lists together in a matrix and *symmetrize* them in order to account for the symmetricity of the friendship relationship as modeled within Facebook: i.e. if A is friend with B, then it should be that B is friend with A too. And also if A is not friend with B, then it should be that B is not friend with A too. It is worth noting that this matrix (the partial friendship matrix), has three possible values in each entry: 0 (we know the two people in question are not friends), 1 (we know the two people in question are friends), and ? (we don't know whether they are friends or not). While performing symmetrization, a 0 or a 1 in entry *(i,j)* will reflect to a 0 or a 1 in *(j,i)*, even if there was a ? originally in that entry. This matrix is *very sparse*, as should be obvious (i.e. the number of 0 or 1 is very small as compared to the number of ?); in our case, we had an 340787 x 340787 matrix, where we had approx 0.001% entries with 1, 0.765% entries with 0, and the rest 99.234%

were unknown. This matrix we later also use for deriving the social information needed for our recognition algorithms Alg2 and Alg3.

***Specifics – Results***

Below you can see plots of the recognition results for the three algorithms: Alg1 w/o social info, and Alg2 and Alg3 with social info. In each of figures 5, 6 and 7, there are two curves: one corresponding to the correct recognition percentage as a function of training set size, and one corresponding to participation in the top-10 ranking subset of the classifier, again as a function of training set size. It is obvious that the latter curve should always be above the former. Two linear fits are also presented above the curves. Correct recognition is in practice useful for a fully automated recognition system (AR); while top-10 participation can be useful for an operator-assisted semi-automated recognition system (SAR), where the top-10 list is presented to an operator for selection. The figures follow below, followed by comments on their meaning:

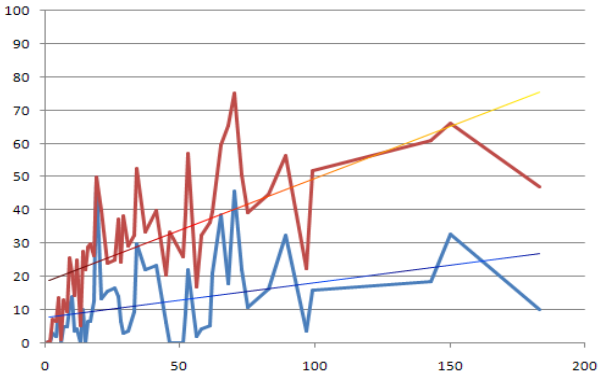


Fig. 5. Alg1 (no social info) Rank-1 and Rank-10 recognition accuracy, as a function of training set size

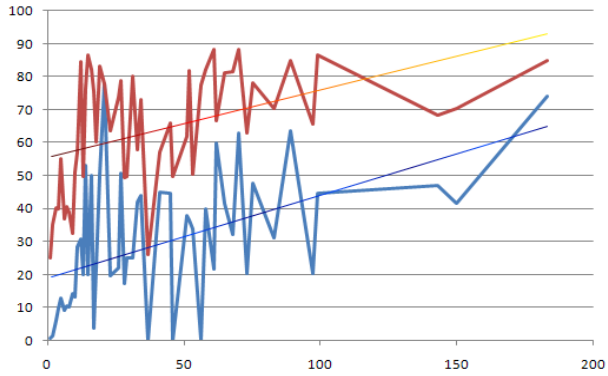


Fig. 6. Alg2 (social info, single label known) Rank-1 and Rank-10 recognition accuracy, as a function of training set size

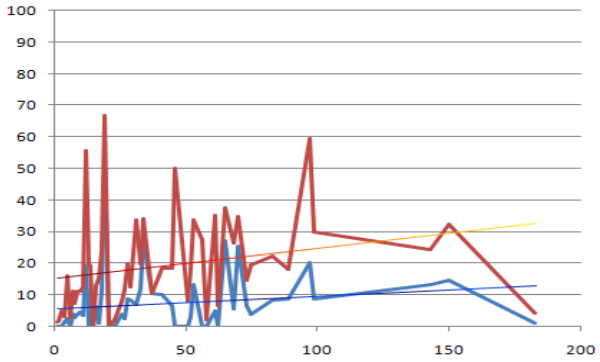


Fig. 7. Alg3 (social info, no labels known, simple algorithm) Rank-1 and Rank-10 recognition accuracy, as a function of training set size

The *first conclusion* to be reached by the figures is that clearly there is a significant increase in recognition performance through the utilization of social information (compare figure 5 to figure 6 for example). In practice, without social info, classifiers made from training sets of size 1-50 or so were totally unusable for both AR as well as SAR; and only remotely helpful in the case of SAR in the case of larger sets (figure 5). However, with social info, one can start using SAR even with training sets of 10 or so, and can definitely use SAR with bigger sets and AR becomes useful with sets over 50. In quantitative terms, across all training set sizes, the rank-1 percentage is 11.5% with Alg1 and 20.3% with Alg2, while the rank-10 percentage grows from 30% to 52.4% (almost a two-fold increase). If we restrict ourselves to only those training sets that have more than four photos, then rank-1 grows from 14.5% to 30%, and rank-10 from 38.5% to 64.4%.

The *second conclusion* to be reached is that although social information can really help, by comparing figure 7 with figure 6 or 5, it becomes clear that a reliable seed is required for this to take place. Alg3 (a very simple algorithm, multiple extensions of which exist) just picks a face at random, calculates recognition scores for it, and chooses the identity that has the highest score as its true identity, and then seeds Alg2 from this. However, if the seed is unreliable, then the social-context-driven boost cannot be so simply utilized. For Alg3, rank-1 and rank-10 percentages are of the order of 4.5% and 12.8% on average; which is even worse than Alg1. Things do not change with larger training sets, too. How can this be improved? Many variations exist, that will be explored in the future including utilizing metrics of overall match across all faces after an initial identity choice for the first face has been made, in order to be able to make alternative choices if the overall match is low. Many other variants exist.

Finally, there is a *third observation* which is very important. The total testing time without utilizing social info is on the order of 23secs per face without parallelization. With social info, through the option of hard-restriction of the hypothesis space, this moves down to 4secs, i.e. a six-fold decrease, quite important in real-time scenarios.

Thus, in short: social information helped us achieve a two-

fold increase in rank-1 and rank-10 accuracies, and has turned unusable results into usable ones. However, one should be very careful when seeding; an unreliable seed can revert the above situation, and a more complicated algorithm than Alg3 has to be used if no seed exists. Finally, social info can also enable a seven-fold speedup.

### *B. Synergies between camera-derived and Facebook-derived pictures towards Face Recognition*

Given that there are two possible sources of photos for our robot (its own camera as well as Facebook-derived photos), we considered the question of the utility of complementing photos from one source with photos from the other for the creation of a better training set. An initial transferability matrix that was calculated in [2] showed that: it does not seem to be worth complementing camera photos with Facebook photos in order to recognize camera photos (drop of 0-1% in recognition accuracy when complementing as compared to camera photos only); and it is certainly not worth to complement Facebook photos with camera photos in order to recognize Facebook photos (drop of 1-4% as compared to Facebook only). Thus, transferability seems bad, and complementing seems useless: one is better off training with camera to test on camera, and training with Facebook to test on Facebook. Or at least, it seems so, for medium-sized training sets (30 photos or more, as is the case in [2]).

However, the question arises: is this also true when one only has very few photos from a source, and wants to test in photos of the same source? Would it be worth complementing the very small native training set with its complement from the other source in this case? The answer can be seen in Figures 8 and 9.

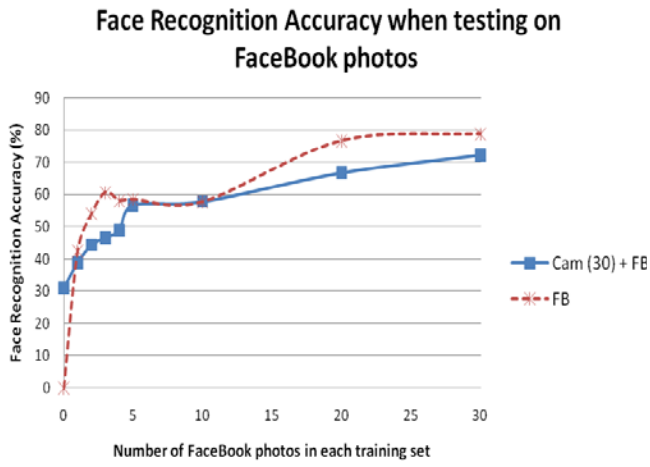


Fig. 8. Complementing across sources: camera photos complementing Facebook photos for the case of small training sets

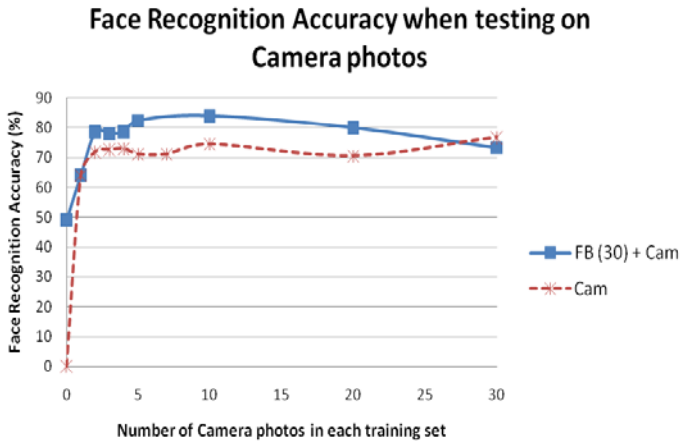


Fig. 9. Complementing across sources: Facebook photos complementing camera photos for the case of small training sets

From the graphs contained in the above figures, it should be clear that there is utility to the above case of complementing across sources: if one only has less than 10 camera pictures at hand, then he is better off training after augmenting his training set with some Facebook pictures, even though one might want to test on other camera pictures again in the future.

This is thus a clear synergy between camera- and Facebook-derived pictures toward face recognition. If one has limited (or no) camera shots for a person, Facebook pictures can help towards better recognition, as illustrated by the above results.

### *C. Interactive Robots assisting online social networking*

Finally, let us consider our third example of mutual benefits. Sarah the Facebot, has the ability not only to access social information and pictures from FaceBook, but also to deposit and publish such information and pictures on her own FaceBook profile, as well as spread information through online chat. This opens up a wealth of possibilities:

First, it is not only the case anymore that humans only can have profiles within Facebook – but robots can do so too. Also, robots can systematically and purposefully collect and deposit information that other humans might find useful, and communicate messages consistently across Facebook-users and physical interaction partners that might not be on Facebook. Furthermore, multiple networked robots can share social knowledge and empirical experience, essentially creating one ultra-social being with multiple geographically distributed bodies, connecting different circles of friends. Also, it is worth noting that the introduction of the robots within the graph of the existing human social networks, can potentially alter the characteristics of the graphs. For example, the social connectivity patterns and statistics of the robot nodes of the graph are expected not to be necessarily similar to the existing human nodes. Finally, the robots can purposefully alter the connections of the graph, for example by introducing and bring together humans or groups of humans which our worth connecting given a goal.

## IV. Conclusion

This paper has explored the intersection of three areas: interactive robots, face recognition, and online social networking, by presenting and discussing an implemented real-world system that combines all three: the "FaceBots" robot, which is a mobile robot with face recognition, natural language dialogue, as well as mapping capabilities. The robot is also equipped with a social database containing information about the people it interacts with, and is also connected in real-time to the "Facebook" online social networking website, which contains information as well as partially tagged pictures. Our system demonstrates the benefits of this triangle of interconnection: for example, we have discussed how Facebook photos enable better face recognition, how Facebook information can lead to more interesting interactions, how interactive robots enable robot-mediated publishing of photos and information on Facebook, and quite importantly, how robot-acquired or Facebook-acquired social information enables better and faster face recognition.

There are also numerous possibilities for future extension, and various open avenues for investigation: for example, one could explore more complex algorithms for utilization of social information, or one could try to predict how the results presented here would scale for larger or smaller sets. One could also try to invert the problem of recognizing faces by utilizing social information derived from the friendship graphs; and thus, try to reconstruct the friendship graph in case it is partially unknown, on the basis of co-occurrence of recognized faces in photos.

In conclusion, the synergies of this three-area triangle framework were exposed through concrete examples: not only interactive robotics has much to gain by utilizing face recognition and online social networking, but all three areas have a wealth of mutual benefits through the many possibilities of useful integration with each other, some of which were explored in this paper.

## Acknowledgment

We would like to thank Microsoft External Research for providing award seed funding for this project.

## References

[1] Mavridis, N., Datta, C., Emami, S., Tanoto, A., BenAbdelkader, C., Rabie, T. (2009) FaceBots: robots utilizing and publishing social information in Facebook, In Proceedings of the 4th ACM/IEEE international conference on Human-Robot Interaction 2009

[2] Mavridis, N., Emami, S. Datta, C., Kazmi, W., BenAbdelkader, C., Toulis, P., Tanoto, A., Rabie, T. (2009) FaceBots: Steps Towards Enhanced Long-Term Human-Robot Interaction by Utilizing and Publishing Online Social Information, Cornell University e-print archive, 2009

[3] Mitsunaga, N., Miyashita, T., Ishiguro, H., Kogure, K., and Hagita, N. (2006) Robovie-IV: A Communication Robot Interacting with People Daily in an Office, In Proceedings of IROS 2006, p. 5066-5072

[4] Breazeal, C., Emotion and sociable humanoid robots, International Journal of Human-Computer Studies, Volume 59, Issues 1-2, Applications of Affective Computing in Human-Computer Interaction, July 2003, Pages 119-155, ISSN 1071-5819, DOI: 10.1016/S1071-5819(03)00018-1.

[5] Brooks, A. G., Gray, J., Hoffman, G., Lockerd, A., Lee, H., and, Breazeal C. 2004. Robot's play: interactive games with sociable machines. Comput. Entertain. 2, 3 (Jul. 2004), 10-10.

[6] M.A.Salichs; R.Barber; A.M.Khamis; M.Malfaz; J.F.Gorostiza; R.Pacheco; R.Rivas; A.Corrales; E.Delgado. Maggie: A Robotic Platform for Human-Robot Social Interaction. IEEE International Conference on Robotics, Automation and Mechatronics (RAM 2006). Bangkok. Thailand. Jun, 2006.

[7] Ishiguro H.; Ono T.; Imai M.; Maeda T.; Kanda T.; Nakatsu R. Robovie: an interactive humanoid robot. Industrial Robot: An International Journal, Volume 28, Number 6, 2001 , pp. 498-504(7)

[8] Okuno, H. , Nakadai, K. and Kitano, H. (2002) Social Interaction of Humanoid RobotBased on Audio-Visual Tracking, In Proceedings of IEA/AIE 2002, p.140-173

[9] Mavridis, N. (2007) Grounded Situation Models for Situated Conversational Assistants, PhD thesis, Massachusetts Institute of Technology

[10] Bennewitz, M., Faber, F., Dominik, J., Schreiber, M. and Behnke, S. (2005) Multimodal Conversation between a Humanoid Robot and Multiple Persons. In Proc. MCHI ws at AAAI05

[11] Michelson, J., and Ortiz, J. (2006) Auto-tagging the Facebook, at: http: //www.stanford.edu/class/cs229/proj2006/MichelsonOrtizAutoTaggingTheFacebook.pdf

[12] Torralba, A. (2003). Contextual Priming for Object Detection. Int. J. Comput. Vision 53, 2 (Jul. 2003), 169-191

[13] Hoiem, D., Efros, A. A., and Hebert, M. (2006). Putting Objects in Perspective. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (June 17 - 22, 2006). CVPR. IEEE Computer Society, Washington, DC, 2137-2144

[14] Stone, Z., Zickler, T. and Darrell, T. (2008). Autotagging Facebook: Social network context improves photo annotation. In Proc. First IEEE Workshop on Internet Vision, June 2008

[15] Kanda, T., Hirano, T., Eaton, D., Ishiguro, H. (2004) Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. In Human-Computer Interaction, 19(1&2):61-84

[16] Ahn, L., Liu, R. and Blum, M. (2006) Peekaboom: A Game for Locating Objects in Images. In ACM Conference on Human Factors in Computing Systems CHI 2006. p. 55-64

[17] Pioneer 2/PeopleBot Operations Manual, ActivMedia Robotics, 44 Concord St., Peterborough NH, 03458

[18] Point Grey Research, Inc.,Vancouver, British Columbia, Canada, web site: http://www.ptgrey.com

[19] Directed Perception, Inc., 890C Cowan Road, Burlingame, CA 94010, web site: http://www.dperception.com

[20] The Internet Communications Engine, ZeroC Inc., web site: http://www.zeroc.com/ice.html

[21] P. Viola and M. Jones. Robust real-time object detection. International Journal of Computer Vision, 57(2):137-154, 2004

[22] D. Bolme, R. Beveridge, M. Teixeira and B. Draper (2003) The CSU Face Identification Evaluation System: Its Purpose, Features and Structure, International Conference on Vision Systems, pp 304-311, Graz, Austria, April 1-3

[23] A. V. Nefian and M. H. Hayes, Hidden markov models for face detection and recognition, in International Conference on Image Processing, 1998.

[24] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel. Sphinx-4: A flexible open source framework for speech recognition. Technical Report TR-2004-139, Sun Microsystems Laboratories, 2004

[25] ActivMedia Robotics, Aria reference manual, Technical Report (1.1.10), 2002.