



A simple yet effective solution to improve the accuracy and efficiency of image search engines can be in semantic annotation of images, i.e. marking the image content with descriptive keywords (labels) or tags. If the image collection is extensively annotated, techniques such as faceted search will help users to filter down a collection and to show potential targets for browsing [3].

There are two methods available to obtain precise image descriptions:

- (i) Automated annotation using computer vision techniques
- (ii) Manual annotation.

Let us first look at the automated methods. Current computer vision algorithms try to extract meaning by analysing the visual content (shape, colour, texture etc.) of an image. However, such approaches have found limited success in specialized settings and are yet to match the performance of humans in image recognition and understanding [4, 5]. The primary reason is for this pitfall is the 'semantic gap' between the low level visual features and high level semantic details.

Humans on the other hand, have little difficulty in describing images, however they find annotation task boring and laborious. Therefore, recently, there have been number of attempts in luring humans to do the laborious task of image annotation. Notable examples are games and social tagging. Although, existing methods are quite successful, these methods in their present form appear inadequate for obtaining annotations of superior quality.

In this paper, as a solution to the above problem, we propose a novel image annotation framework based on the theory of Emergent Semantics. We describe the motivation and need for such approach. A working prototype of the proposed approach is presented in the form of entertaining online game, called GoFish. GoFish is a web variation of standard Go Fish [6], a popular playing card game. People play the game for fun and entertainment. And as a side effect of playing the game, images get tagged (annotated). We carried a small user study to test the viability of the design. Result of the study is encouraging.

## 2. RELATED WORK

Humans have mostly avoided tagging despite its benefits in terms of recall and retrieval. Researchers therefore, tried different approaches to entice humans in annotating images. Three most prominent of these approaches are (Refer Figure 2):

- (i) Embed tagging in social activity
- (ii) Provide monetary incentives
- (iii) Design special purpose games.



Figure 2: Effective methods to entice human participation

### 2.1. Social activity

An easy way to turn something boring into something that is interesting is to make it social. For example, tagging images may be boring to individuals, but together in a group (with friends and relatives) it can be fun and interesting. Ludicrop Inc. [7] realized very early the inherent social nature of humans and designed a social tagging system, Flickr [8] around it. Flickr is a popular image hosting and sharing website. Its popularity has been fuelled by its organization tools: mainly tagging. Tagging allows users to attach set of textual labels (tags) freely to images and browse with them. The benefit of using Flickr is simple social platform where people can come together, share photos, and tag them collaboratively. Currently, Flickr claims to host more than 4 billion images [9].

However, in Flickr, tagging is a choice. As a result, there exist many unlabelled images in Flickr from uninterested users. Since tagging is not forced on to the users, the users, who are trying to increase their exposure, will only tag images. Beyond the restricted community of friends, there is little reason for an average user to tag any image properly.

### 2.2. Monetary incentives

Images can also be annotated by paying humans the money in return. The concept was introduced with Amazon Mechanical Turk (AMT) [10] which coordinates workers and developers in solving human intelligence tasks like tagging, for a small payment in return. Some recently launched search engines such as TagCow [11] utilize AMT. They pay \$1.20 per hour to participants tagging images. However, current image search engines do not have an alternate source of generating revenue like Advertisements on image search pages. Therefore, the strategy of paying humans to tag images is not

well justified from the business point of view. Moreover, getting unbiased, precise description for images from unknown contributors is also a problem of concern.

### 2.3. Special purpose games

Each year, people around the world spend billions of hours playing computer games. The field of *Human Computation* [12] started with the aim of channelling all this time and energy into useful work such as annotating images. Sometimes people like to think and be challenged; sometimes it is just for pastime. Online games are thus seductive methods for encouraging people to participate in a collaborative work such as tagging. Such games constitute a general mechanism for using brain power to solve large scale problems. In fact, designing such a game is much like designing an algorithm—it must be proven correct and its efficiency must be analysed [12]. People play such games for entertainment, and not because they want to voluntarily tag images. Existing human algorithmic games designed for image annotations are: ESP game [13] and Phetch game [14].

Let us look at them individually.

#### 2.3.1. Extra Sensory Perception (ESP) game

ESP is the first human computation game, proposed by Luis von Ahn et al in 2004 [13]. It has been hugely successful and popular among the users. Millions of image tags have been collected via playing the game, and even after a many years of its launch, people are still interested in playing the game.

The ESP game play is very simple. In this game, two randomly paired players try to agree on labels for single image. On each successful match (i.e. both players choose the same tag for the image), both players score points. The word or tag on which the two players agree, then becomes the Taboo word for that particular image. In the next round of game with the same image, players cannot use the Taboo words to describe the image. They must agree on some new label (tag) for the image in order to score points in the game. Authors argue that if the image has generated an extensive list of Taboo words (words that player cannot use) and players are unable to agree upon new label and preferring to pass the image, then image can be considered as completely labeled [13].

However, problem with the ESP game is that it encourages users to assign obvious labels to images, primarily because, doing so can easily lead to an agreement with the partner. For example, let us assume the image in Figure 3 is shown to players in ESP game.



**Figure 3:** Which tags will result in easy agreement with the partner in ESP game?

Even if one or both the players know the name of the person in the image (that is, Gary Oldman in this case) they will not tag the image as 'Gary Oldman'. Since the pairing is done at random, the player does not know whether his partner can also recognize the man in the image as Gary Oldman (Refer Figure 3). Therefore, their obvious guesses or tags will be limited to: 'man' and 'spectacles'. These tags are not wrong in any sense, but the question arises whether one has to rely on humans to obtain them.

Moreover, ESP game gives players an easy option of passing on difficult image and difficulty is kept up to the user to decide. Therefore, in order to score more points in the game, player will prefer to pass the image with taboo words, rather than thinking over to extract new tags. This fact can also be seen in their game statistics as only 1023 of 293760 images have five or more labels [13].

#### 2.3.2. Phetch game

Phetch [14] game is aimed at fetching natural language descriptions for images, to help blind people in navigating images. In this multiplayer game, one player (Describer) describes the image to other players (Seekers) and they try to find the image from a search engine for given description. On success, all players get points.

However, in order to enjoy the game play it is "must" for either Describer or Seeker or both of them to possess some knowledge about the given image. If Describer doesn't describe the image properly to the Seekers, then it is hard for seekers to find desired image. For example, consider the image shown in Figure 4 (similar image is also shown in original paper [14]).



**Figure 4:** To search for this image, players must know the names of the celebrities.

If this image is given to describer, he must know that the man and the woman shown in the image are Justin Timberlake and Janet Jackson respectively. A general query describing the scene like “two singers, a man and woman in a concert when man ripped a piece of woman’s shirt” is not sufficient to search with. Moreover, for such a description, seeker should also be knowledgeable enough to replace the man and woman with Justin Timberlake and Janet Jackson Respectively. This assumption is too strict and will therefore fail to attract global audience.

### 3. MOTIVATION

On reviewing the existing literature, we found some fundamental problems in the way humans tag images. For example, most of the times, the person who is tagging an image completely ignores the possibility that a searcher for the image may not know what he is looking for or may not be able to recollect what he wants. Seeing and saying may have meaning to one observer, but the same visual experience may not have the same meaning to another observer.

To illustrate, a person while tagging follows his own interpretation of image and tag accordingly. He may not be aware of other possible or complete interpretations of the image, or can simply ignore them. For him, ‘the meaning of image is what struck to his eyes’. In effect, the tagged image remains accessible only to him and to people who also interpret the image similarly. Any other person, who wishes to find the same image but queries differently, will not possibly find the image due to mismatch in interpretation with the owner or tagger.

The person, who is tagging an image, must realize that the image itself does not have any meaning. It is merely a rectangular shape with amorphous blotches of various sizes and colours. While looking at image, we interpret and compare the blotches to objects or situations we encountered before. The cumulative of all visual clues in an image gives us the ability to constitute context and meaning.

To understand it better, let us follow a thought experiment and observe our tagging tendency.

#### 3.1. Fallacies of misplaced concreteness

Let us first understand how we generally describe an image. The way we describe the image is as much function of “how we see” as it is function of “how we think” or more appropriately “how we are made to think” [15]. To illustrate, let us describe the image we see in Figure 5.



**Figure 5:** How will you describe this image?

Figure 5 shows a familiar pattern of oval shape and green colour. Therefore, we reply as “It is an Apple”. Now, consider the same image, within a group of other images as shown in Figure 6 and try describing it.



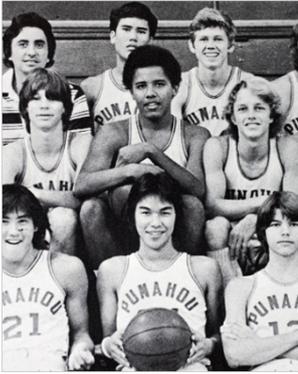
**Figure 6:** Is ‘apple’ still a good tag for the first image?

Our discriminating mind can see that all images are of Apple, yet we know, the first one is different from the rest. We therefore, look for the features that separate this image from the rest of the images and describe the first image in Figure 5 as “It is an apple fruit.” If we progress in the same way and assemble the same image with other images of Apple fruit, our description becomes even more precise and we say “It is a green apple fruit, partially eaten from left.”

This small exercise in the prequel shows that, ‘we do not always say what we see’. We all saw and knew from the beginning, the features present in the image, but we never felt the need to express it completely. We described the image with an abstract notion of an ‘Apple’. The problem with manual annotation is our tendency to oversimplify things. We tend to get lost in what Alfred North Whitehead called as, ‘Fallacies of misplaced concreteness’ [16]. If we do not describe the image precisely (i.e. communicate the complete cognitive experience through language), the image under view remains hidden in the crowd of similar images and needs cumbersome browsing for retrieval.

### 3.2. Crippled viewer syndrome

Greisdorf and O'Connor in their book [15] coined the term *Crippled viewer syndrome* which refers to the cognitive disconnect a viewer experiences on seeing an unknown image. Without the background knowledge necessary to interpret the image, a viewer can describe the image only in terms of the signs it contains. Therefore, the true (intended) meaning of the image is often does not get communicated in the annotation. For example, let us look at the image in Figure 7. If we do not know that the person in the middle is 'Barack Obama', then our description would probably be limited to 'Basketball team', 'black kid among white kids', 'group photo' etc.



**Figure 7:** Without the necessary background knowledge, most people will tag this image as 'basketball team'

These problems with manual annotation exist because, to most individuals, applying word descriptors to a textual document and to an image appear to be the same sort of activity. However, they cannot be the same activity, since describing the document text is an extraction process, and there are (usually) no words to extract from an image (photo, painting, etc.) Creating a structured collection of images based on words requires an underlying framework that connects the collection to its viewers through purposeful communications [15], which we present in the next section.

## 4. EMERGENT SEMANTICS THEORY

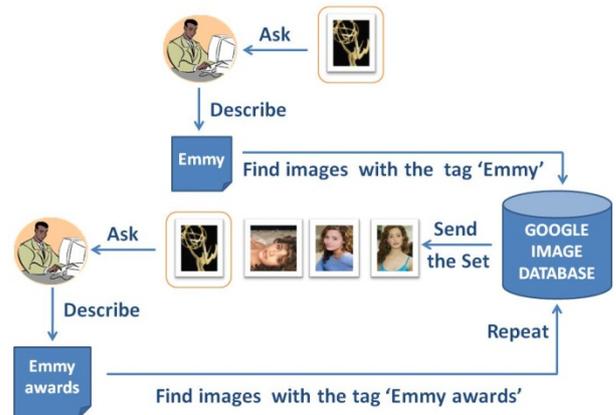
A solution to the annotation problem lies in the theory of *Emergent Semantics* [17]. According to this theory, image in general does not have meaning, but the meaning emerges from the interaction with the user and by placing the image in the context of other images. The small exercise in the prequel is a proof for the same. When the image of the apple is placed in the context of other images as shown in Figure 5, we are able to describe (or made to describe) the image more precisely.

To elaborate more, emergent semantics theory [17] reveals that:

- **Meaning of the image is contextual**  
It depends on the particular condition under which the annotation is done and particular user that is annotating the image.
- **Meaning of the image is differential**  
Meaning of the image can be made manifest by differentiation between an image which possesses that meaning and image which does not. Further, Meaning of the image can also emerge by association between different images that share that meaning.
- **Meaning of the image is grounded in action**  
Meaning of the image can also be established from the user actions when the image is presented to her.

### 4.1 Emergent semantics approach to image annotation

Using the emergent semantics theory, we present a novel approach for annotation of images (Refer Figure 8). Our approach is a recursive way of extracting new meanings of the image by repeatedly placing the image in the context of other similarly described images.



**Figure 8:** Emergent Semantic Based Framework for Image Annotation

#### 4.1.1. The procedure

The steps are as follows:

Present user with an image  $A$  to describe.

Get the description  $D_0$  for the image  $A$ .

Find all images from the database that corresponds to the given description  $D_0$ .

Present the original image  $A$  along with images found in the step 3 and ask user to describe the original image again with respect to other shown images.

Get the new description  $D_1$  for the image  $A$ .

Repeat the steps 3 to 5 using the new description  $D_1$ .

The process stops when the user is unable to differentiate between the images and add a new description (tag) to the image. At the end, each image will have a rich set of  $n$  image tags,  $D = \{D_0, D_1, \dots, D_n\}$ .

## 4.2. Benefits of this approach

### 4.2.1. Productivity:

If an image  $A$  receives a description or tag  $D$ , then the tag  $D$  not only describes the image  $A$ , but it also tells us that tag  $D$  differentiates the image  $A$  from the rest of the presented images. These accompanying images therefore, will not have the same tag  $D$ . As a result, annotation is done faster and on the complete set of presented images.

### 4.2.2. Features ranking:

With every new round, the image  $A$  receives a new description  $D_i$ . In the first round people describe the most striking feature of the image (For example 'apple' for image in Figure 5). In all subsequent rounds; next most striking features about the image are introduced (For example, 'fruit' 'green' 'eaten' for the image in Figure 5). This hierarchical way of tagging helps in maintaining a ranking of the received tags.

However, success of the above approach depends upon active human participation, which can only happen if humans find this task engaging and entertaining.

We therefore present an intelligent system design for tagging images using the above mentioned emergent semantics approach in the form of a game. Our technique like previously proposed games [18], is not dependent upon computer vision techniques, but on people's existing perceptual abilities and desire to be entertained. However, unlike previously proposed games, rather than creating a new game for image annotation, we actually transform an existing popular game into a game with a purpose of image annotation. We first describe the original game and later we discuss how to transform it to suit our purpose.

## 5. GOFISH: TRANSFORMING A POPULAR GAME

Go Fish is popular card game, played among two to five players with a deck of 52 playing cards. One of the players is chosen as a *Dealer*, who first shuffles the cards and distributes them equally among all the players including him. Objective of the game is to win most Books of cards where, a Book is a collection of four cards of same rank. For example, four kings, four aces, etc. Figure 9 shows interface of an online version of GoFish game [19].



Figure 9: A screenshot of an online Go Fish card game

Player to the left of the *Dealer* starts the game. He asks any one of the players, for a card of specific rank and from a specific suit (Hearts, spades, clubs and diamonds). For example, "John, Do you have '6 of hearts'?" However, in order to ask, the player himself must have at least one card of the same rank, i.e. '6' in this case. If *John* has the requested card, he has to give it to the player who asked for it. Whenever, the request for the card is successfully fulfilled, the same player continues asking for other cards. But if the player addressed i.e. *John* in this case, does not have the requested card, he says "Go Fish!" It means, player who asked for the card, loses his turn for asking and now *John* can start asking for cards. Once the player collects all the four cards of specific rank to complete one Book, he shows them to all and keeps them face down on the table. The game proceeds in the same manner until all the thirteen books of cards are won. The player with most number of books is declared as the "Winner".

### Transformation

Let us see, how we can transform this game into a game with a purpose of efficient tagging of images. As a first thought, the transformation seems easy. Just replace the playing cards with images. The catch here is, whenever a player asks for a card, he has to describe it in plain text. If we capture all such description, it will solve the problem of describing the images. However, we must ensure that cheating is minimal and generated descriptions are accurate. We describe below the modified version of the GoFish.

### 5.1. GoFish: Our proposed game

GoFish is a turn based game played among four players. One of the players is chosen as *Narrator* and others are *Seekers*. The deck of the cards is a collection of eight images those we wish to tag. The replica of the entire deck is always visible at the bottom of the game screen as shown in Figure 10.



Figure 10: Screenshot of a GoFish Game in action

Figure 10 shows a snapshot of GoFish game window. The four players are seen at the four corners of Figure 10.

Narrator starts the game by shuffling and then distributing the cards equally among all the players (including him). Therefore, each player holds two cards (they are marked with orange bubble. Refer Figure 10). The players do not know other players' cards. The objective of the game is to win (collect) all the cards.

### 5.1.1 Game play

Narrator gets the first chance to ask for a card. In order to ask, Narrator must enter proper description of that card which can be approved by Seekers. He describes the card (that he wants to ask) in the plain text and sends the description to all the Seekers. All the Seekers on their respective turn try to identify the card that matches the received description (all the cards are visible at the bottom of the screen, refer Figure 10). Seeker scores points for finding the correct card. If majority of Seekers is able to find the correct card, Narrator gets points for the valid description. However, if none of the Seekers is able to find the correct card, then Narrator is penalized for incorrect description. Narrator will lose his turn of asking after two such penalties.

Once the card has an agreed description, Narrator picks any one player and asks him for the described card. If the player has that card, he has to give it to the Narrator. Narrator then continues asking for more cards. However, if the player does not have the requested card, he says "Go Fish!" and becomes the Narrator. He can then ask for the cards. Present Narrator takes his position as Seeker. The game continues in the same way till one player wins all the eight cards, he is then declared as "Winner".

### 5.1.2 Strategy

GoFish maintains a scoreboard which lists top scorers and players who won maximum number of games on the current day, week and till date (all time winners). We present below strategy for winning the game and scoring good points.

#### 5.1.2.1 Winning a game:

To win the game, a player needs to collect all the cards. At the start of the game, player does not know which player has which card (Probability of correct guessing is 1/3). For a player to improve his chances, a good strategy is to pay attention to who seeks which card He can then capture those cards in the next turn if he can remember whom to ask.

#### 5.1.2.2 Scoring High points:

Narrator scores points for every card he correctly describes, while Seeker gets points for every correct card he finds. If card is correctly described every time and all the Seekers are able to find the card then everybody gets equal points and none gets chance to become Narrator for the next game. To beat other players in scoring, Narrator can opt to give description to which minimal number of Seekers agree (Minimum number of Seekers will be able to find the correct card). Narrator cannot give wrong description, to which no Seeker will agree. Therefore, a better strategy would be to describe the image in more specific details, hoping that not all the Seekers knew about it. It is a gamble, but worth taking. Similarly, for Seeker to score high he should look to find the correct card each time and follow the above strategy if he gets chance to become Narrator after wards.

## 5.2. Description quality

A *proper* description is *correct* if it makes sense with respect to the image and *complete* if it gives enough information about its content. The description becomes *superior* if it conveys beyond what can be seen from the image.

### 5.2.1. Accuracy

We argue that descriptions generated by playing GoFish will always be accurate. We list following points in support of it.

- All the players are randomly grouped from all the players online to avoid colluding.
- Narrator cannot give description that does not correspond to any image (irrelevant) or more than one image (incomplete). In both the cases, he will lose points as Seekers may not be able to find the correct card.
- For an easy agreement with the Seekers, Narrator might want to describe the position of the card as discriminating factor, for example, "First image", "second from right". We make sure that no match (agreement) is possible by

- If Seeker select wrong image for given description, he gets negative points. Since the players are randomly grouped, the probability that all the seekers choose same image which is different from the one Narrator has picked is low.

### 5.2.2. Completeness

We expect that Narrator will describe the image with only features that separates the image from the rest of the images. These features may not be sufficient to describe the image completely. We therefore follow emergent semantics theory discussed earlier and group the image with other similarly described images in a new game instance of GoFish. Now, Narrator of the new game cannot give the same description as before, as it will now correspond to two or more images (The previous description cannot separate the image from the rest of the images). Therefore, Narrator must explain the image further to score points in the game. An image can be said to be completely described if Narrator can no longer able to distinguish the image from the rest of accompanying images and asks for replacements.

### 5.2.3 Superiority

We call a description Superior, if it conveys beyond what is conveyed (what we can see) in the image. Players who are playing the game are viewers, not the originator of the images. Getting superior description by player (viewer of the image) compared to the originator of the image is difficult. The viewer may not know the meaning of what he is seeing, yet he knows what he is looking at. We intend to help him by giving supportive images which are similarly described to obtain the desired background knowledge. Even then a player may not give superior description for an image. To make him do that, we introduce the strategy of beating other player in points as we explained in Section 5.1.2.2. Player (Narrator) can attempt to give superior description, hoping that not all, only one Seeker knows that the description holds (makes sense) for the selected image.

## 6. IMPLEMENTATION AND USER STUDY

GoFish is implemented in Adobe Flash [20] and Smart fox server [21] is used for socket connections. Upon completion of the game, server records all activities of the player in the database for future analysis. Currently, GoFish is in beta stage. GoFish is made available within the

university campus for testing. We present below the results of preliminary study conducted.

### 6.1. Objective

The objective of the user study was to know answers for following questions:

- Is playing GoFish fun?
- Do people want to play GoFish regularly?
- Does the game generate valid tags?
- What is the quality of generated tags?

### 6.2. Participants

A total of 30 players played the game over the period of 2 weeks. All participants were students from the university campus with their age in the range of 19 to 28 with 4 female students and 26 male students. To avoid the cold start, monetary incentives were provided for winners and image set was selected from the top 20 results of 10 popular search queries from India using Google Zeitgeist [22]. Table 1 shows the 10 popular search queries used to evaluate GoFish.

**Table 1:** Search queries used to evaluate GoFish

Amrita Rao	Apple
Cars	Cricket
Harry Potter	Hrithik Roshan
Lord Ganesh	Sachin Tendulkar
Shah Rukh Khan	The Dark Knight

### 6.3. Procedure

A web based prototype as mentioned before is created and its URL was mailed to all the participants. A tutorial was provided along with the mail explaining the rules of the game and it was also kept on the game URL page as shown in Figure 11.

Any player can start a new game or join an existing game as shown in game list in the right panel of Figure 11. However, a game does not start unless four players (including the one who started it) have joined and are ready to play.



Figure 11: Welcome page of GoFish game

Once that happens, a new game will start, indicating it to all the four players who joined. The player who starts the game will become the first Narrator and remaining three players become Seekers.

## 6.4 Results

### 6.4.1. Efficiency and accuracy

Each game lasted for roughly 14 minutes. 60% of the players played the game more than once, while 8 players played the game for four or more times. Total number of games played was 33 that generated 231 descriptions for the 150 images. We analyzed the quality of the generated tags and found that 78% of the received descriptions were specific to the image, 12% general and 10% superior. Table 2 shows examples of these descriptions for four search queries.

Table 2: Example tags received with GoFish

Search query	Descriptions		
	General	Specific	Superior
Sachin Tendulkar	 Face	 century	 Rohan Gavaskar
Apple	 Green	 Sliced apple	 Music
Cars	 Orange car	 Fancy car	 Lightning McQueen
Harry Potter	 Draco	 Movie Poster	 Red Carpet

Note that, general description is the basic level description of the image while specific description speaks about the striking feature of the image and superior description conveys beyond what is seen in the image. For example in Table 2, under Apple category, 'Green' is general tag for an image of a green apple while 'sliced apple' is specific tag and 'music' is a superior tag for an image of iPhone.

### 6.4.2. Enjoyability

In correspondence with the objective 1 and 2, upon completion of the game, we requested users to answer a set of questionnaire providing feedback

about playing the game and to write down any specific comments they have. This step was optional and 24 out of 30 users participated in it. First question was: whether playing GoFish was fun? A 5 point scale was given to users to record their answer with 5 being extremely enjoyable and 1 being least enjoyable or boring. GoFish received on average a score of 3.2 on the 5 point scale. Second question was whether they would like to play GoFish again? And answer was to be given in binary form (yes/no). Most of the players (75%) said they would love to play the game again.

In the comments section of the questionnaire, some users asked a single player version of the game, while few users suggested its integration with social networking sites and introduction of time gap between the moves.

We believe above results are good indicator of the viability of the design. As a future work, we are concentrating on the comments and the suggestions and trying to build a single player version of the game with its integration to Facebook [23].

## 7. DISCUSSION

GoFish is a game that tests not only player's ability to distinguish among images but also his memory and above all his luck. To score high points, player needs to describe image correctly, while to win the card, he needs good luck and concentration in the game. A factor like randomness which comes with luck was missing from the earlier games like ESP and Phetch [13, 14].

One criticism on GoFish is that it is not simple like ESP. However, GoFish is adopted from an existing popular game and we kept the game play nearly same as the original. Therefore we believe players who loved original Go Fish card game, will also appreciate this design for its novelty. Our user study indicates that most users easily understood the rules and liked the game with only 20% needed hands on demo.

With GoFish, we introduce competitive factor in the Game with a Purpose design [18]. While earlier games like ESP and Phetch are collaborative in nature. GoFish is aimed at receiving more search specific tags. Although, we can also obtain good tags with the existing games, but it will require an alternation of the game, which we believe, will take away the fun.

Unlike earlier proposed games, GoFish game can also be played among friends and families and not limited to randomly grouped players. A person can first upload all his photos to our game site, and then invite his friends for playing the game. In that

way he not only gets his photos tagged, but also gets fun of playing with his friends.

## 8. CONCLUSION

Until recently, research in image annotations has largely been centered on development of effective techniques such as games to lure human into annotation. However, less focus was given on the quality of the resultant annotation. We discussed the problems with extant methods and presented a different perspective on annotation using semantic theory of images. We introduced an intelligent annotation scheme GoFish. We explained the design with potential benefits. At the end, we gave a preliminary user study to show the viability of the approach and the game. Although the game is not released to public, we hope in the near future, our game will help annotate majority of the images.

## 9. ACKNOWLEDGMENTS

We would like to thank anonymous reviewers, C-STAR lab-mates and the participants for their valuable comments. We also thank Rutuja Khot for proofreading the paper.

## 9. REFERENCES

- [1] Google Image search. <http://www.images.google.com/>.
- [2] Bing Image search <http://www.bing.com/images/>.
- [3] White, R.W., Kules, B., Drucker, S.M., and schraefel, m.c. (2006). Supporting Exploratory Search, Introduction to Special Section of Communications of the ACM, Vol. 49, 4(2006), 36-39.
- [4] Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R. Content-Based Image Retrieval at the End of the Early Years. IEEE Trans. Pattern Anal. Mach. Intell. 22(2000), 1349-1380.
- [5] Datta, R., Joshi, D., Li, J., and Wang, J.Z. 2008. Image retrieval: Ideas, influences, and trends of the new age. ACM Computing surveys, 40(2008), 83-85.
- [6] Go Fish. [http://en.wikipedia.org/wiki/Go\\_Fish/](http://en.wikipedia.org/wiki/Go_Fish/).
- [7] Jefferson, G. Flickr of idea on a gaming project led to photo website. USA Today. [http://www.usatoday.com/tech/products/2006-02-27-flickr\\_x.htm](http://www.usatoday.com/tech/products/2006-02-27-flickr_x.htm).
- [8] Flickr. <http://www.flickr.com>.
- [9] 4 Billion Photos on Flickr from Flickr Blog. <http://blog.flickr.net/en/2009/10/12/4000000000/>
- [10] Amazon Mechanical Turk <http://aws.amazon.com/mturk/>.
- [11] Image Recognition Problem Finally Solved: Let's Pay People To Tag Photos. <http://www.techcrunch.com/2008/03/29/image-recognition-problem-finally-solved-lets-pay-people-to-tag-photos/>
- [12] Von Ahn, L. 2005 Human Computation. Doctoral Thesis. UMI Order Number: AAI3205378., Carnegie Mellon University.
- [13] Von Ahn, L., and Dabbish, L. Labeling images with a computer game. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'04, (2004), 319-326.
- [14] von Ahn, L., Ginosar, S., Kedia, M., Liu, R., and Blum, M. 2006. Improving accessibility of the web with a computer game. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '06. ACM, New York, NY, 79-82.
- [15] O'Connor, B. C., and Griesdorf, H. F. 2007. Structures of Image Collections: From Chauvet-Pont-d'Arc to Flickr, Libraries Unlimited.
- [16] Fallacy of misplaced concreteness. [http://en.wikipedia.org/wiki/Fallacy\\_of\\_misplaced\\_concreteness/](http://en.wikipedia.org/wiki/Fallacy_of_misplaced_concreteness/).
- [17] Santini, S., Gupta, A., and Jain, R. Emergent semantics through interaction in image databases. IEEE Trans. Knowl. Data Engg. 13(2001), 337-251.
- [18] Von Ahn, L., and Dabbish, L. Designing games with a purpose. Commn. Of ACM, 51(2008), 58-67.
- [19] Online Game Gofish. [http://www.littlebrowniebakers.com/games/gofish/gofish\\_play.htm](http://www.littlebrowniebakers.com/games/gofish/gofish_play.htm).
- [20] Adobe Flash 8 <http://www.adobe.com/products/flash/>.
- [21] Smart Fox Server <http://www.smartfoxserver.com/>
- [22] Google Zeitgeist <http://www.google.com/press/zeitgeist/>.
- [23] Facebook. <http://www.facebook.com/>.