Vol.2010-SLDM-144 No.47
Vol.2010-EMB-16 No.47
Vol.2010-MBL-53 No.47
Vol.2010-UBI-25 No.47
2010/3/27

写真と書き込みの共有による協調体験を強化するエージェント

森元　俊成[†1]　角　康之[†2]　西田豊明[†2]

本研究では写真撮影と手書きメモの共有に基づいて興味や気づきの共有を行う PhotoChat 上にユーザが気づいていない情報を付加しながら，PhotoChat 上のチャットに参加するエージェントの導入を試みる．その際，エージェントは構造的に整理された情報を，ユーザの書き込んだキーワード，位置，書き込みの有無といったユーザの反応に従って提示するようにした．エージェントの導入により，ユーザが注目している対象について気づきを得ることができるだけでなく，コンテンツやサービスの提供者側の立場では，押し付けがましくなくユーザーの満足度の高い状態で写真型コンテンツの提供やサービスへの誘導ができると考えられる．

Agent that Enhances Ourdoor Experiences by Pseudo Chats based on Photos and Notes

Toshinari Morimoto ,[†1] Yasuyuki Sumi [†2]
and Toyoaki Nishida [†2]

We developed an agent which shows information which users have not noticed and participates conversation in PhotoChat, which enables users share their interests and discoveries based on photos and notes. The agent shows information which is sorted out according to users' reaction, including keywords they write, places they write in and whether or not they write. By introducing an agent, there are merits for both users and providers of content or services. Users can discover something new about things they look at, and providers can introduce their content or services to users without irritating them.

†1 京都大学 工学部
Kyoto University, Faculty of Engineering
†2 京都大学 情報学研究科
Kyoto University, Graduate School of Informatics

1. はじめに

人々は日々，様々な活動を通して，グループ内で体験を共有しながらコミュニケーションを行っている．そこで,「見ているもの」に対して書き込みを行うことで気軽に興味や情報への気づきを共有することを可能とする手段として，写真撮影と手書きメモの共有に基づいたツール PhotoChat[1)] を開発してきた．本研究では，その上にユーザが気づいていない情報を付加して，PhotoChat 上のチャットに参加するエージェントの導入を試みる．ただし雑多でまとまりのない情報や，ユーザの反応を無視した情報提示は，かえってユーザの不満を招くと考えたため，構造的に整理された情報をユーザの反応にしたがって提示するようにした．今回は PhotoChat を用いながらグループが観光を行っている状況において，エージェントが会話に参加することを想定している．

2. 関連研究と本研究の位置づけ

見ているものを他のユーザと共有して，それに対して会話するというアプローチを取るものとして岡田ら[2)] の DigitalEE II がある．このシステムでは，現場にいる学習者と遠隔にいる学習支援者がカメラ映像と音声でつながっている．現場の学習者がカメラ型携帯情報端末を持ち，興味のある対象にカメラを向けると遠隔の支援者がカメラ越しに，本やコンピュータ上の関連情報を調べて音声で伝えるシステムである．現場の学習者にとっては遠隔の支援者は現実強化の役割を果たしているといえる．本研究では，この遠隔の支援者にあたる役割をエージェント化した．DigitalEE のように，学習者の対話的な質問に対して正確に回答するものではなく，見ているものに対する関連情報をエージェント主導で提供し，新たな気づきを与える，新たな疑問を発見させることを目的としている．また DigitalEE は映像と音声対話を使っているのに対し，我々は写真とそれに対する手書き文字による書き込みという，手軽で直感的な手段を使っているので，興味の対象を切り出し，その上に持続的に会話を積み重ねていくことが可能である．

本研究では，ユーザ同士の会話に参加しながら情報提供をするエージェントの実現を目指しているが，質問応答エージェントの実現を目指しているものではない。またユーザ主導の会話にエージェントが参加することは大変難しいため，エージェント主導で話題提供を行うことを目指す．そこでエージェントが提供する話題が断片的なものにならないために，話題

Vol.2010-SLDM-144 No.47
Vol.2010-EMB-16 No.47
Vol.2010-MBL-53 No.47
Vol.2010-UBI-25 No.47
2010/3/27

同士の関連性や発展性のネットワークを持ちユーザの反応に応じて話題展開を行うこととした．このようなストーリーに基づいた既存研究には，コミックダイアリ[3]がある．コミックダイアリでは，コミュニティメンバの体験に基づいて，漫画のストーリーを提示することでコミュニケーションを支援している．

屋外で携帯情報端末や GPS 機能付き携帯電話を利用して観光や課外学習を支援するシステムは，これまでにも多く提案されてきた (例えば Cheverst[5] や垂水ら[4])．しかしこれらの多くは，前もって用意された情報をユーザ各自の状況合わせて提示するものであり，基本的にシステムとユーザ個人の間の対話に基づいている．それに対し，我々のシステムは，ユーザ同士によるシステム上の会話からグループとしての興味の焦点を読み取り，それに合わせて情報提示をすることを試みる．

3. 写真と書き込みの共有による仮想会話に参加するエージェント

3.1 仮想会話を強化するエージェント

3.1.1 理想的なシステム動作イメージ

図 1 は，本研究で目指しているシステムの理想的な動作イメージを示している．図 1 は，仮想会話への参加者は端末を用いて体験の共有を行っており動物園でキリンを見学しているとする．そこで，エージェントは例えば飼育員といった一般的な見学者とは違った立場の人間による経験に基づいた知識を提供し，ユーザーに知識や気づきといったものを提供することを想定している．例えば図 1 のように親のキリンしか公開されておらず，飼育員などの限られた人間にしか親子の姿が見られない場合もある．そこで親子のキリンの姿の写真や見学のポイントといった情報を提示することで，一般のユーザーでも飼育員といった特別な立場の人間の体験に基づいた知識や気づきが得られより豊かな体験を得ることができると考えられる．

3.1.2 本研究の特徴

本研究での特徴は，人とエージェント間の会話において言語情報だけでなく，ユーザーの特定の写真への書き込みや，写真上の特定エリアへの書き込みといった点を利用しながら構造的に整理された情報を提示することで，会話に参加するところにある．

3.1.3 エージェント導入による効果

エージェントが仮想会話に参加することで注目している対象について他の人の知識や体験に基づいて気づきを得ることができるだけでなく，コンテンツやサービスの提供者側の立場では，押し付けがましくなく，ユーザーの満足度の高い状態で写真型コンテンツの提供やサービスへの誘導ができると考えられる．

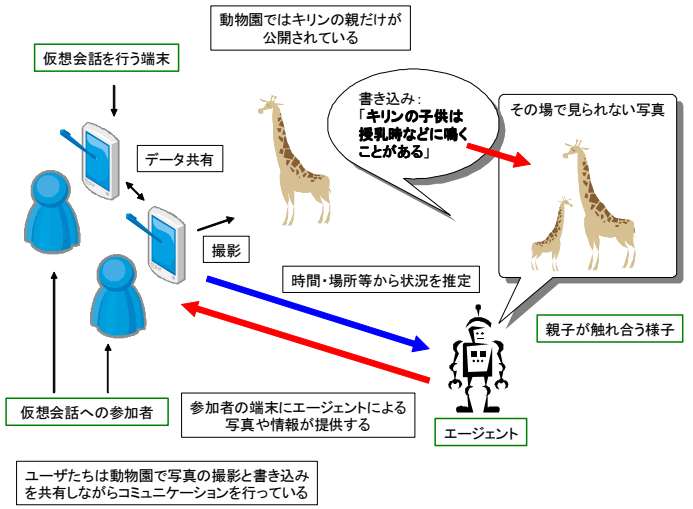

図 1 システムの理想的な動作イメージ

3.2 PhotoChat:写真と書き込みの共有に基づいたコミュニケーション支援ツール

本研究では，屋外で行動を共にしながら観光や課外学習に参加するメンバーを想定する．そこで，写真撮影とペンによるメモ書きといった直感的な手段を電子的に共有する PhotoChat[1] と呼ばれるコミュニケーション支援ツールを基盤システムとし，その上に，PhotoChat ユーザの一人として振る舞うエージェントを実装することを試みる．PhotoChat[1] は，写真撮影とその写真上での手書き文字を複数のユーザー間で共有できるようにしたものである．図 2 に PhotoChat の画面の例と使用中の写真を示す．ユーザーがカメラで写真を撮影すると PhotoChat で使用されているネットワークを介して，接続されているすべての端末で同様の写真が共有される．また，手書きの文字についてもすべての端末上で共有される．さらに，PhotoChat には写真上で，他の写真へのハイパーリンクを貼り付ける機能が存在する．ハイパーリンクの機能を利用することで，写真同士の関連を表現することもできる．本研究では，仮想会話の手段として PhotoChat を用い，PhotoChat 上で行われる仮想会話に参加するエージェントを実装した．

Vol.2010-SLDM-144 No.47
Vol.2010-EMB-16 No.47
Vol.2010-MBL-53 No.47
Vol.2010-UBI-25 No.47
2010/3/27

図 2 PhotoChat の使用画面

3.3 エージェント構築の設計方針

エージェントの開発にあたって以下のような要件が必要であると考えられる．

- 時間・空間の認識
  - 時刻情報から季節，時間帯といった概念に変換する
  - GPS 情報からエリアのスポット情報，ランドマークに変換する
- ユーザ質の撮影，書き込み，閲覧といった振る舞いのモニタリング
  - ユーザ達の撮影，書き込み，閲覧の時間パターンからユーザ達の興味のスポットを認識する
  - 書き込みの認識
  - 書き込み作業の時空間セグメンテーション
  - 文字や典型的なストロークパターンの認識
- 情報の提示を行う能力
  - ユーザ達の注目している対象の周辺の写真を提示したり，書き込みを行ったりする
  - ユーザが閲覧，書き込みで応答した場合に，さらに内容を深めて情報提示をする
- コンテンツ
  - 写真，図などの図的データに基づいた話題の情報
  - 話題間の関連性を管理するための背景知識

4. 実　　装

この章では実装に関する詳細について述べる．エージェントの動作は，ユーザーのコンテキストを理解する「状況認識プロセス」と，新たな話題を提供したり写真上での会話に参加したりする「会話参加プロセス」からなる．状況認識プロセスでは，住所や時間に関する情報を取得し，得られたデータを特定のランドマーク等に変換したり，ユーザの写真上での書き込みの状況を認識するプロセスである．状況認識プロセスでの動作は節 4.1 で説明を行う．会話参加プロセスでは，状況認識プロセスで得られたユーザの周辺のランドマークに関わる写真を提示したり，ユーザの書き込みに応じてエージェントが書き込みを行ったり関わりのある写真を提示したりするプロセスである．それについては節 4.2 で説明を行う．

4.1 状況認識プロセス

エージェントの状況認識プロセスにおける処理は主に 3 つからなる．

- 住所と時間から特定のランドマーク等と結びつける処理 (4.1.1)
- ユーザの書き込みの状況を認識処理
  - ユーザの各写真に対して書き込みの有無の認識
  - ユーザがどのようなキーワードを書き込んでいるか認識 (4.1.2)
  - ユーザの各写真上で注目している領域を認識 (4.1.3)

である．状況認識プロセスを表す図を図 3 に示す．図ではユーザがとある住所に 12 時に居たとすると，その時間帯に開演している付近の動物園の情報を結びつけを行うことを示している．また図のように，写真上の「きれいだね」という書き込みがあったとすると，そこから「きれい」というキーワードを認識したりクジャクの頭を囲っている円から，ユーザがクジャクの頭部に興味を示していることを認識したりする処理を行う．

4.1.1 住所と時間からランドマーク等を結びつける処理

エージェントが初めの話題投稿のきっかけとして，住所と時間情報に基づいて写真投稿を行う．エージェント用の端末は一定時間ごとに (本実装では 12 秒ごとに設定) 現在地の座標を取得し，直ちに逆ジオコーディング用のデータベースを用いて住所情報へと変換する．次に住所と時間情報に応じて特定の写真を選び出す処理を行う．以下に住所と時間情報と投稿される写真の関係を表したデータベースの例を示す．なお以後は「住所と時間情報のデータベース」と呼ぶこととする．

```
Geo=京都府, 京都市 左京区, 田中門前町,103 Photo=chionji.jpg
Geo=京都府, 京都市 左京区, 田中門前町,103 Date=*,*,15 Photo=tedukuri.jpg
```

Vol.2010-SLDM-144 No.47
Vol.2010-EMB-16 No.47
Vol.2010-MBL-53 No.47
Vol.2010-UBI-25 No.47
2010/3/27

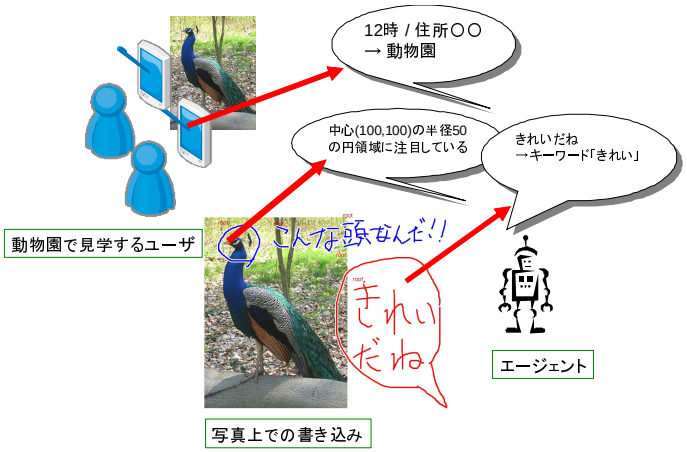

図 3 状況認識プロセス

`Geo=京都府,京都市 左京区,吉田泉殿町,1 Hour=11,12,13,14,15 Photo=museum.jpg`
住所と時間情報のデータベースは，各行が住所と時間情報および結びつけられるランドマーク等の写真の対応を表す．例えば 1 行目では，Geo=の後に指定された場所である「京都府京都市左京区田中門前町 103」の付近に訪れた場合に Photo=の後に指定された chionj.jpg という写真 (知恩寺) が関わりのあるランドマークの写真として選び出される．また住所だけでなく，時間情報の制約を加えることもでき，2 行目の Date=*,*,15 は「日付が 15 日ならば」ということを意味している．それによって毎月１５日に知恩寺で行われている「手作り市」というイベントがあることを，その日たまたま近くを通りかかった人に知らせることができるようになる．カンマで区切られた 3 つの数字は先頭から年，月，日を意味しており，アスタリスクは任意の場合を表している．また 3 行目は博物館の写真と結びつくことを表しているが，Hour=11,12,13,14,15 という指定によって 11 時台〜15 時台でないと結びつきが起きないことを表している．それによって開館時間に合わせて写真の提供を行うことが可能となる．

4.1.2 書き込んだキーワードを認識する処理

PhotoChat ではユーザーの書き込みは点 (座標) の集合を繋いだストロークの集合で表現されている．ユーザの書き込んだキーワードを認識するために，ストロークの集合をまずテキストに変換する必要がある．テキストに変換するにはストロークの集合を文字認識モジュールに渡してその結果を得ることで行う．文字認識モジュールは既存のものを用いる．ただし，文字認識モジュールは横書きで 1 行分を前提にしているため，それに対応するために各ユーザのストロークをいくつかのグループに分ける．そのため時間的・距離的に一定範囲内にある同一ユーザーによるストローク集合を同一のグループに分類する．なおストロークのグループ化のアルゴリズムは伊藤[1] の実装で既に行われていたものである．ただし本実装の際には元々の設定値を少し変更し 5sec，50pixel 以内であることをグループ化の基準とした．上記の値を設定した理由は，1 つのストロークグループが最大でも 1 行分になるような設定である程度満足がいくものであったからである．しかし，同一単語の文字列が別のストロークグループに分けられることもあったので，同一ユーザーによるものならば，変換されたテキストは後ですべて結合することとした．例えば「美しい」と書いた場合に「美」と「しい」に相当するストロークが別のグループに分類されてしまうといった場合である．

次に，ストロークの集合から文字認識モジュールを用いて変換されたテキストは，タグ変換テーブルを用いてタグに変換される．表記が異なっていても統一的に扱えるようにするためである．例えば「なぜ」,「どうして」のような同じ意味の単語はどちらも「why」というタグに変換される．以下はタグ変換テーブルの例である．

```
なぜ=why
どうして=why
きれい=beautiful
美しい=beautiful
```

各行でイコール記号「=」で区切られたうちの左側がテキスト形式で，右側がそれに対応するタグである．文字認識によって得られた文字列が，各行の左側の文字列と同じものを含んでいれば，右辺のタグを追加する.「美しい」という文字列からは「beautiful」というタグに変換される．ただし，本エージェントでは言語情報にはあまり頼らないものを目指しているため，言語による状況認識は副次的なものである．

4.1.3 写真上の注目している領域を抽出する処理

ユーザの興味を推定するために，本エージェントは写真上の特定領域にも着目を行う．ユーザは写真上で注目した特定のオブジェクトを丸で囲んだり，また矢印で指したりというような言語情報以外のパターンを用いることもある．エージェントはユーザの写真上での書き込みで丸や矢印を検知した場合にどの領域を指しているか計算を行う．丸で囲われた場合

Vol.2010-SLDM-144 No.47
Vol.2010-EMB-16 No.47
Vol.2010-MBL-53 No.47
Vol.2010-UBI-25 No.47
2010/3/27

は，中心の座標と半径を計算し，矢印で指し示された場合は，その先の座標を求める．便宜的に矢印の場合は，半径 1 の円領域として処理する．こうして得られた領域は以後，「ユーザが注目している領域」と呼ぶこととする．ただし現段階では，丸や矢印の検出には先程と同様に文字認識モジュールで代用している．本来は図形は文字ではなく，図形として認識される必要があるため，今後は図形の認識では改善が必要である．

4.2 会話参加プロセス

エージェントは状況認識プロセスで得られた，ユーザの状況をもとに会話を行う．会話参加プロセスでの会話へのアプローチは主に次の 2 つからなる．

- 住所や時間をもとにした新たな話題の提供
- ユーザの書き込みをもとにした書き込みや写真提供

である．

4.2.1 住所や時間をもとにした新たな話題の提供

状況認識プロセスでは，一定時間ごとに住所を取得し，住所と時間の情報を用いてデータベースを参照し，特定のランドマーク (の写真) に変換する作業を行った．会話参加プロセスでは，そこで選び出された写真をユーザに提供する．ただし，選び出された写真に関して，住所が番地まで正確であることは期待できないため，住所に関しては番地以外が一致していれば提供される写真の候補に追加をすることにした．そして取得された住所との番地の誤差の大きさに応じて，投稿される写真をいくつかのグループに分けて，誤差の小さいグループをユーザに提供するようにした．

4.2.2 ユーザの書き込みをもとにした書き込みや写真提供

エージェントによって提供された写真上では，エージェントはユーザの書き込みに応じて，書き込みを行ったり，新たな写真を提示して貼りつけたりする．エージェントは自身の提供した写真に関しては背景となる知識を持っている．その背景となる知識を，実装上では知識データベースと呼ぶこととする．状況認識プロセスにおいて得られたユーザの書いたキーワードや，特定の領域に関する注目を表す情報，またユーザの写真に対する書き込みの有無といった情報をもとに知識データベースを参照し，情報提示を行うことで会話に参加する．図 4 にエージェントが書き込みや，写真の貼付けを行う例を示す． 知識データベースは以下のように記述されている．この知識データベースは図 4 のクジャクの画像に対応する知識データベースである．

```
ID=001 Each=beautiful All=!ugly Answer=羽を広げた時 Photo=k2.jpg
ID=002 X=90 Y=80 Radius=1 Photo=kujaku_head.jpg
```

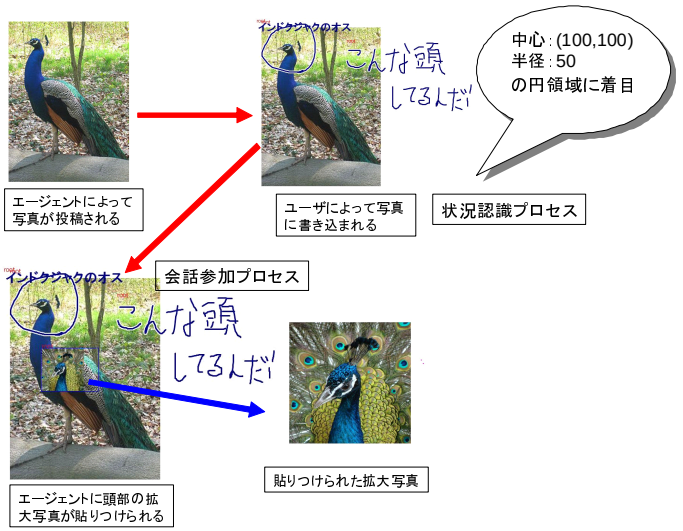

図 4 会話参加プロセス

知識データベースの各行には，書き込みの内容・貼り付ける写真と，それらが書き込まれる (貼りつけられる) ための条件式が記述されている．各行には，「パラメーター名=値」の組み合わせが「タブ」で区切られて記入されている．各パラメーターとそれの表すものは以下の通りである．

(1) 「ID=」の後には各行を識別するための ID を書く．
(2) 「Each=」の後には，ある同一のユーザーが書いた書き込みに対応するタグの集合に関する条件を書く．
(3) 「All=」の後には，すべてのユーザーの書いたすべての書き込みに対応するタグの集合に関する条件を書く．
(4) 「Answer=」の後には，出力の条件が満たされた場合にエージェントによって書き込まれる文字列を書く．
(5) 「Photo=」の後には，出力の条件が満たされた場合にエージェントによってハイパーリンクとして貼りつけられる写真のファイル名を書く．
(6) 「Interest=」の後には，他のユーザーが書き込んだかどうかに関する条件を書く．0 を指定すれば，どのユーザーも書き込んでいない場合を表し，1 を指定すれば他の

Vol.2010-SLDM-144 No.47
Vol.2010-EMB-16 No.47
Vol.2010-MBL-53 No.47
Vol.2010-UBI-25 No.47
2010/3/27

ユーザーが書き込んだ場合を表す．

(7) 「X=」「Y=」「Radius=」の後には，写真上の特定のオブジェクトが存在する領域を円領域で表すために，x 座標と y 座標と半径を書く．

(8) 「Same=」の後には，同じ回答を行っている行の ID を書く．これは，出力するための条件が異なっていても，出力される内容が同じ場合に，重複して同じものを出力しないための存在する．

知識データベースの例の 1 行目では「Each=beautiful」,「All=!ugly」という部分が条件式にあたる部分である．なお「ugly」の直前の「!」の記号は「ugly」が存在しない場合を意味している．状況認識プロセスでは，ユーザがどのような書き込みを行ったかを認識して，それをタグという形式に変換した．そのとき「美しい」「きれい」といった文字列を「beautiful」というタグに変換して,「醜い」といった文字列が「ugly」というタグに変換される．2 行目の条件式は，あるユーザーが「美しい」と書き込み，また全体で誰も「醜い」と書き込まれていなければそれに対応する書き込みと写真の貼りつけが行われることを表している．

知識データベースの例の 2 行目は図 4 で示した例に相当する．条件式として「X=90」「Y=80」「Radius=」1 が指定されている．これは，写真上の (x,y)=(90,80) の位置に半径 1 の大きさの円領域上にオブジェクトが存在していることを意味している．そのオブジェクトが，状況認識プロセスにおいて得られた「ユーザの注目している領域」内に存在する，または逆に「ユーザーの注目している領域」内にオブジェクトが存在する場合に条件を満たす．図 4 では，状況認識プロセスにおいて中心 (100,100) で半径 50 の円領域をユーザが着目していると認識しており，その領域内に，中心 (90,80) で半径 1 の円の領域が含まれていたため，会話参加プロセスで，kujaku_head.jpg(クジャクの頭部拡大写真) の貼りつけが行われたことを表している．なお，このときエージェントは「こんな頭してるんだ！」という言語情報には触れずに，このような動作を行っている．なお，このときエージェントは「こんな頭してるんだ！」という書き込みの言語的意味を解釈しているわけではない．頭の部分に連想付けされた話題を知識データベースに持っており，ユーザの書き込みに反応して，クジャクの頭部拡大写真を提示しているだけである．深い意味処理を行わずに，写真上の注目エリアに他の話題を関連づけるという単純な知識構造を利用するという割り切りが，本研究の特徴であると共に，エージェントの対話能力の限界でもある．

5. 動 作 例

図 5 では，金閣寺周辺を観光しているユーザがいて，その住所と時間帯の状況からエージェントが金閣寺におけるワンシーンとして舞妓が金閣寺を訪れたときの写真を提示していることを表している．次に図 6 では，図 5 に引き続いて，エージェントが投稿した写真に対してユーザが異なった書き込みをしている．図 6 は，同じ写真でもユーザーの注目している領域に応じて異なった写真が貼り付けられる場合を表している．上の段のように，ユーザーが舞妓に着目すれば別の舞妓の写真を提示する．下の段のように，金閣寺の上部のオブジェクトに着目していれば，それに関連した情報を提示することを示している．このようにユーザの着目している領域に応じて選択的に話題を深化させるように動作する． 今回の動作例

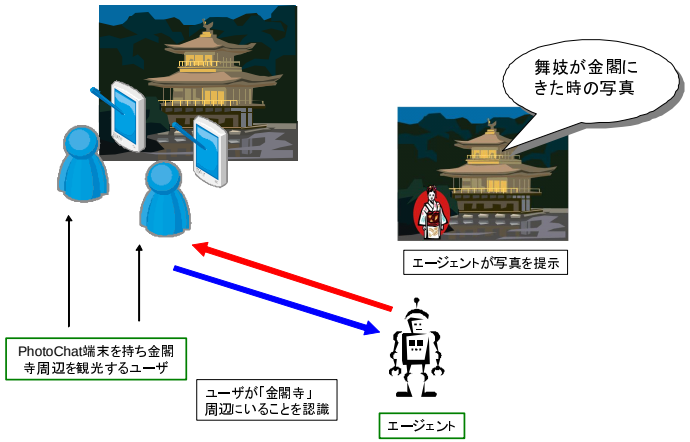

図 5 動作例 (1)

で背景となっている知識は図 7 に示すようになっている．舞妓の写っている領域上には，そこに舞妓が写っていることを知識として持っており，同様に金閣寺の上部の領域上には，そこに鳳凰が写っていることを知識として持っている．また，特定のキーワードに反応して，「足利義満」の画像を提示するということも知識として保持している．新しく投稿された鳳凰，足利義満，舞妓の写真にも背景知識が存在するため，そこから次々と関連のある話題をユーザの反応に応じて深化させることができる．今後はそれらをより一般的なストーリー展開のための知識ネットワークを定式化していくつもりである．

Vol.2010-SLDM-144 No.47
Vol.2010-EMB-16 No.47
Vol.2010-MBL-53 No.47
Vol.2010-UBI-25 No.47
2010/3/27

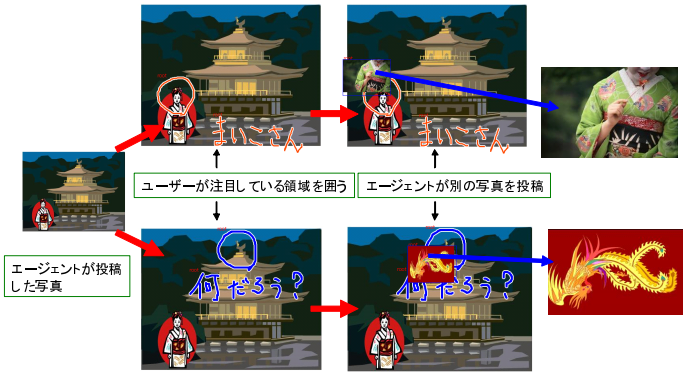

図 6 動作例 (2)

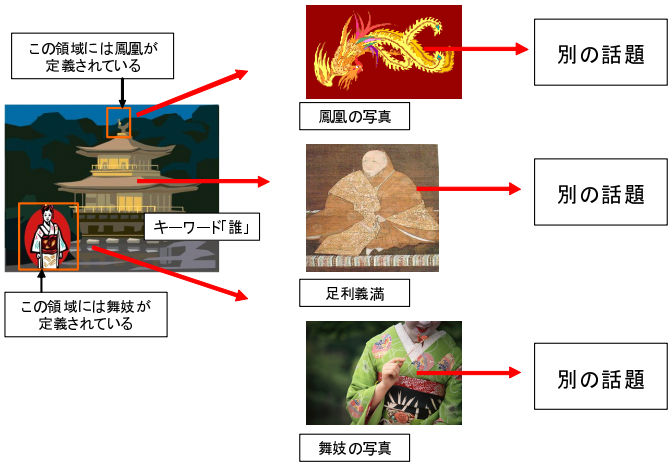

図 7 背景知識の例

6. おわりに

本研究では，PhotoChat 上でユーザの反応に従って構造的に整理された情報を提示するエージェントの実装を行った．現時点ではタイミングや提示する情報量などについては考慮していないため，今後はエージェントが写真を提示したり書き込んだりするタイミングや量も考慮に入れる必要がある．またコンテンツの作成において，一般的なストーリー展開のために，話題同士の関連を定式化された知識ネットワークで表現していく必要がある．

参 考 文 献

1) 角 康之, 伊藤 惇, 西田 豊明：PhotoChat：写真と書き込みの共有によるコミュニケーション支援システム, 情報処理学会論文誌, Vol.49, No.6, pp.1993-2003 (2008)
2) 岡田 昌也, 山田 暁通, 吉田 瑞紀, 垂水 浩幸, 粥川 隆信, 守屋和幸：現実・仮想経験拡張型システム DigitalEE II による協調型環境学習, 情報処理学会論文誌, Vol45, No.1 pp.229-243 (2004)
3) Yasuyuki Sumi, Ryuuki Sakamoto, Keiko Nakao, Kenji Mase : ComicDiary: Representing individual experiences in a comics style, Proceedings of Ubicomp 2002 (Springer LNCS2498), pp.16-32 (2002)
4) 垂水浩幸、鶴身悠子、横尾佳余、西本昇司、松原和也、林勇輔、原田泰、楠房子、水久保勇記、吉田誠、金尚泰：携帯電話向け共有仮想空間による観光案内システムの公開実験、情報処理学会論文誌、Vol.48, No.1, pp.1882-7764 (2007)
5) Keith Cheverst and Nigel Davies and Keith Mitchell and Adrian Friday and Christos Efstratiou : Developing a context-aware electronic tourist guide: Some issues and experiences, Proceedings of CHI 2000 (ACM), pp.17-24 (2000)