ネットワーク情報のAR可視化システムの提案

大沼 信也[1,a)]　阿部 亨[2,b)]　菅沼 拓夫[2,c)]

概要：情報通信技術（ICT）の普及により，専門的知識を持たない一般ユーザであっても，ネットワーク機器の各種設定・操作に携わらざるを得ない状況が増えてきている．このような一般ユーザの作業を支援するために，本稿では，Augmented Reality (AR) 技術を用い，対象であるネットワーク機器の実際の映像へ作業に必要な情報を重ねて可視化するシステムを提案する．提案システムでは，ユーザが持つ携帯端末で，対象であるネットワーク機器を撮影した際に，各ポートの状態等のネットワーク情報を映像中の適切な箇所に随時表示することで，一般ユーザが行う各種作業の支援を図る．

1. はじめに

情報通信技術（ICT）の普及により，専門的知識を持たない一般ユーザでも，ネットワーク機器の各種設定・操作に携わらざるを得ない状況が増えてきている．このような一般ユーザの作業を効果的に支援するためには，ネットワーク機器の設定・操作に必要な各種の情報をユーザへ適切に提供するシステムが有効であると考えられる．これを実現するために，本研究では，Augmented Reality (AR) 技術を用い，対象であるネットワーク機器の実際の映像へ，作業に必要な情報を重ねて可視化するシステムの開発を目指している．これまでにも，ネットワーク管理に AR 技術を応用する研究がなされているが，それらの多くは，ネットワーク情報を遠隔地のユーザへ単に提示する程度に止まっている [1]．これに対し，本稿では，ネットワーク機器を対象に，ユーザが持つ携帯端末でこれを撮影した際，ネットワーク情報を映像中の適切な箇所に（例えば，各ポートの状態を当該ポートの近傍に）随時表示することで，ユーザが行う作業の支援を図るシステムを提案する．これによって，従来のネットワークマネージャとは異なり，現場の作業に沿った情報の提示が可能となり，作業の効率化が期待される．

携帯端末で撮影したネットワーク機器の映像中の適切な箇所に各種の情報を重ねて可視化するためには，映像中でのネットワーク機器の位置・姿勢（3 次元空間での携帯端末とネットワーク機器との相対的位置関係) を正確に決定する必要がある．対象の位置・姿勢を映像から推定するために，形状が既知のマーカーを利用する手法が広く用いられている．この手法は，対象の所定の箇所に所定の向きでマーカーを事前に貼付し，これを撮影した映像中でのマーカーの見え方から，マーカー（および対象）の位置・姿勢を推定するものであり，AR マーカーと呼ばれる特定のマーカーを利用し，映像を撮影するカメラとマーカーとの相対的位置関係を推定する ARToolKit などが開発されている [2]．一般に，スイッチ等のネットワーク機器を複数台運用する環境では，個々を管理するためのラベルを各機器に貼付する場合が多い．したがって，この管理用ラベルをマーカーとして利用できれば，ラベルを貼付した機器の位置・姿勢も映像から推定することが可能となる．しかし，人手で機器に貼付したラベルには，所定の箇所・向きからのずれが必ず生じるため，ラベルがマーカーとして利用でき，その位置・姿勢が映像から正確に推定されたとしても，そこから得られた機器の位置・姿勢には推定誤差が含まれる．一方，対象の全体の形状を表現した CAD モデルをあらかじめ用意しておき，映像中の対象の見え方と CAD モデルとを比較することで，対象の位置・姿勢を推定する手法がある [3]．この手法では，マーカーを対象へ貼付する必要がないものの，ネットワーク機器がラックに収納されている環境では，対象の一部（フロントパネル等）の見え方しか利用できないため，位置・姿勢の正確な推定は困難である．また，位置・姿勢が既知である対象のテンプレート画像と入力された映像とで特徴点のマッチングを行うことにより，マーカーを用いずに，入力映像中での対象の位置・姿勢を推定する手法がある [5]．この手法では，複数の

1 東北大学大学院情報科学研究科
Graduate School of Information Sciences, Tohoku University, Aoba, Sendai, Miyagi 980–8577, Japan
2 東北大学サイバーサイエンスセンター
Cyberscience Center, Tohoku University, Aoba, Sendai, Miyagi 980–8577, Japan
a) onuma@ci.cc.tohoku.ac.jp
b) beto@cc.tohoku.ac.jp
c) suganuma@cc.tohoku.ac.jp

©2015 Information Processing Society of Japan

図 1　想定する対象ネットワーク機器の例

図 2　管理用ラベル（マーカー）の例

ポートを持つネットワーク機器のように，見掛けが類似した箇所が多数存在する対象の場合，特徴点の対応を一意に決定することが難しく，位置・姿勢の正確な推定は困難である．

そこで，提案システムでは，マーカーとして利用可能な管理用ラベルが貼付されたネットワーク機器を対象に，マーカーを用いた手法と特徴点のマッチングによる手法を組み合わせることで，携帯端末とネットワーク機器との相対的位置関係を正確に推定し，映像中の適切な箇所での情報の可視化を行う．その際，表示の精度はネットワークスイッチのポート半分未満を目標とする．

本稿では，まず，第 2 章で，提案するネットワーク情報の AR 可視化システムの概要を述べる．第 3 章では，映像中でのネットワーク機器の位置・姿勢の推定手法を提案し，第 4 章では，実験により提案手法の有効性を検証する．第 5 章では，まとめと今後の課題を述べる．

2.　ネットワーク情報の AR 可視化システム

本研究では，図 1 に示すように，ラックに収納された複数のネットワークの機器を対象として想定している．各機器には，個々を管理するために，図 2 に示すようなラベルが貼付されているものとし，ラベルを貼付する箇所・向きはあらかじめ定められているとする．ただし，ラベルは人手で機器に貼付されるため，所定の箇所・向きからのずれが生じ得るとする．

本研究で開発を目指すシステムを用いた作業のイメージを図 3 に示す．ユーザが，対象となるネットワーク機器群を自身の携帯端末のカメラで撮影すると，携帯端末は，まず，映像中の管理用ラベルを認識する．ユーザはその中から認識したい対象の管理ラベルを選択する．携帯端末は選択された管理ラベルを基に機器の特定を行い，次に，特定された機器に関連するネットワーク情報（ネットワーク構成，トラフィック，各ポートの状態など）を管理サーバーから取得する．これと並行して，携帯端末は，対象機器の位置・姿勢を映像から推定する．推定された対象機器の位置・姿勢に基づき，携帯端末画面に表示された機器の映像中で適切な箇所を決定し，取得したネットワーク情報を映像に重ねて可視化する．ユーザは，携帯端末の画面を確認しながらネットワーク機器の設定や配線等を行うことができるため，十分な専門的知識を持たない場合でも，容易に作業を進めることが可能になると期待できる．

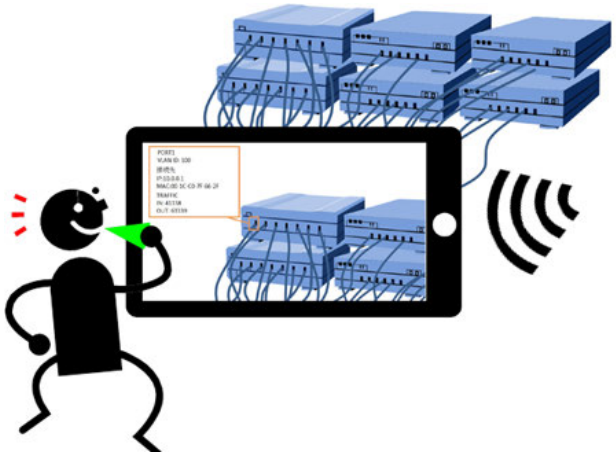


図 3　作業のイメージ

3.　対象ネットワーク機器の位置・姿勢推定

ネットワーク情報の効果的な AR 可視化を行うためには，携帯端末で撮影した映像中で，対象ネットワーク機器の各部位を正確に特定し，情報の適切な表示箇所を決定する必要がある．これを実現するために，提案システムは，以下の手順で，映像中でのネットワーク機器の位置・姿勢の推定を行う．

(1) 映像の歪み補正
(2) 管理用ラベルを利用した大まかな位置・姿勢推定
(3) 特徴点マッチングによる高精度な位置・姿勢推定

3.1　映像の歪み補正

携帯端末のカメラは小型の広角レンズを使用しているため，撮影した映像には大きな歪みが生じている場合が多い．そこでまず，使用端末について事前にカメラキャリブレーションを行うことで得られるパラメータを用いて，映像の歪み補正を行う．以降の処理では歪みの無い画像を扱う．

3.2　管理用ラベルを利用した大まかな位置・姿勢推定

次に，提案システムは，歪みの補正された映像から，ARToolKit [2] を用い，各ネットワーク機器に貼付された

©2015 Information Processing Society of Japan

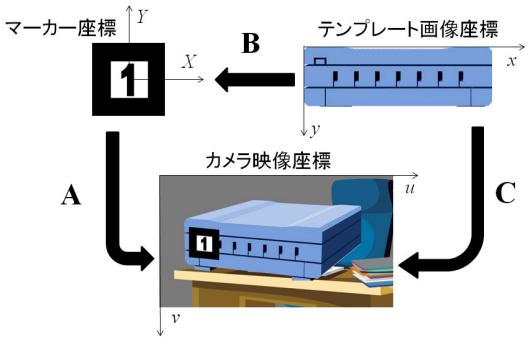


図 4　映像座標系，ラベル座標系，テンプレート画像座標系

図 5　テンプレート画像の例

管理用ラベル（マーカー）の検出と認識を行い，その結果から，映像中でのネットワーク機器の位置・姿勢の大まかな推定を行う．

図 4 に示すように，ラベル座標系と映像座標系を考える．ラベルの面が $Z = 0$ であるとすれば，ラベル座標系の点 (X, Y, Z) $(Z = 0)$ を映像座標系の点 (u, v) へ投影する変換は次式で表される．

$$\begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{1}$$

ARToolKit は，登録されたラベルのパターンとの類似度が最大となる四角形の領域を映像中で決定し，これにより，ラベルが投影された映像中の領域を検出する．また，ラベルの各頂点と検出領域の各頂点の対応を決定することで，式 (1) 中の変換行列 $\boldsymbol{A}$ を推定する．

一方，ネットワーク機器へラベルを貼付する箇所・向きがあらかじめ定められているならば，機器を正面から撮影したテンプレート画像座標系（図 5）の点 (x, y) をラベル座標系の点 (X, Y) へ投影する変換行列 $\boldsymbol{B}$ を事前に求めることができる．ネットワーク機器へラベルを正確に貼付した場合に，テンプレート画像の x 軸，y 軸，原点が，ラベルの X 軸，Y 軸，原点と各々一致するよう座標系を設定すると，(x, y) を (X, Y) へ投影する変換は次式で表される．

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} m & 0 & 0 \\ 0 & -m & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \boldsymbol{B} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2}$$

ここで，m は，テンプレート画像とラベルとのスケール比を表す．

式 (1) と (2) から，テンプレート画像中の点 (x, y) を映像中の点 (u, v) にへ投影する変換は次式で表される．

$$\begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \boldsymbol{AB} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

したがって，ネットワーク機器の各部位の位置をテンプレート画像中で事前に求めておけば，ネットワーク機器の位置・姿勢により，各部位が映像中のどこに投影されるかを式 (3) から推定することができる．

しかし，式 (3) から推定される映像中の位置は，ARToolKit による変換行列 $\boldsymbol{A}$ の推定誤差に加え，ラベルの実際の箇所・向きと変換行列 $\boldsymbol{B}$ のずれ等によっても大きな影響を受ける．そのため，管理用ラベルを用いた対象ネットワーク機器の位置・姿勢の推定のみでは，各ポートの状態を当該ポートの近傍に表示する等，ネットワーク情報を適切な箇所に AR 可視化するに十分な精度が得られないことが予想される．

3.3　特徴点マッチングによる高精度な位置・姿勢推定

映像中での対象の位置・姿勢をより高精度に推定するために，テンプレート画像と入力映像とで特徴点のマッチングを行う手法が考えられる．

テンプレート画像中の点 (x, y) を映像中の点 (u, v) に投影する変換が次式で表されるとする．

$$\begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \boldsymbol{C} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4}$$

テンプレート画像と映像とで特徴点を各々求め，テンプレート画像中の特徴点 (x_n, y_n) と映像中の特徴点 (u_n, v_n) の対応が N 組 $(N \geq 4)$ 得られれば，RANSAC アルゴリズム [7] 等により変換行列 $\boldsymbol{C}$ を決定ができる．その際，テンプレート画像と映像とで正確に対応付けられている特徴的対が多い程，$\boldsymbol{C}$ の推定精度は高くなる．しかし，映像中に同型のネットワーク機器が複数台写っている場合や，ネットワークポートのように見掛けが類似した箇所が多数存在する機器が対象の場合，テンプレート画像と映像とで特徴点の類似度から対応を正確に決定することは困難である．このため，誤って対応付けられた特徴点対が増え，$\boldsymbol{C}$ を正確に求めることが難しい．そこで提案システムでは，管理用ラベルを用いて推定された対象ネットワーク機器の大まかな位置・姿勢に基づき，対応する特徴点を探索する際の範囲を制限することで特徴点マッチングの精度向上を図り，正確な $\boldsymbol{C}$ の推定を実現する．

いま，テンプレート画像座標系の点 $p = (x, y)$ をラベル座標系に投影する場合を考える．管理用ラベルが所定の箇所・向きに貼付されている（ずれが無い）場合の投影位置を $P = (X, Y)$，管理用ラベルがラベル座標系で所定の箇所から (t_x, t_y) ずれている場合の投影位置を $P_t = (X_t, Y_t)$，所定の向きから θ ずれている場合の投影位置を $P_\theta = (X_\theta, Y_\theta)$

©2015 Information Processing Society of Japan

とすると，管理用ラベルのずれにより生じる投影位置のずれは各々以下の通りとなる．

$$|P_t - P| = \sqrt{t_x^2 + t_y^2} \tag{5}$$

$$|P_\theta - P| = \sqrt{2(X^2 + Y^2)(1 - \cos\theta)} \tag{6}$$

管理用ラベルを貼付する際に許容される所定の箇所・向きからのずれが最大 (t_x, t_y)，θ であるとすれば，式 (5) と (6) から，ラベルのずれにより生じる投影位置のずれの最大値 R を次式で求めることができる．

$$R = \sqrt{t_x^2 + t_y^2} + \sqrt{2(X^2 + Y^2)(1 - \cos\theta)} \tag{7}$$

したがって，映像中で抽出された特徴点を (u_i, v_i)，テンプレート画像中で抽出された特徴点を (x_j, y_j) とすると，両者が対応する特徴点であるならば，ラベル座標系で以下の関係を満たすと考えられる．

$$\sqrt{|X_i - X_j|^2 + |Y_i - Y_j|^2} < R \tag{8}$$

ここで，(X_i, Y_i) と (X_j, Y_j) は，次式により，(u_i, v_i) と (x_j, y_j) をラベル座標系に投影した位置である．

$$\begin{bmatrix} w'X_i & w'Y_i & w' \end{bmatrix}^{\mathrm{T}} = \boldsymbol{A}^{-1} \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^{\mathrm{T}} \tag{9}$$

$$\begin{bmatrix} X_j & Y_j & 1 \end{bmatrix}^{\mathrm{T}} = \boldsymbol{B} \begin{bmatrix} x_j & y_j & 1 \end{bmatrix}^{\mathrm{T}} \tag{10}$$

提案システムでは，テンプレート画像と映像とで特徴点マッチングを行う際，式 (8) の条件を探索範囲の制限に用い，この条件を満たすものの中で類似度が最大となるものを対応する特徴点として選択する（ただし，最大の類似度が閾値以下である場合は，対応する特徴点が無いと判断し，変換行列 $\boldsymbol{C}$ の推定には用いない）．これにより，特徴点の誤った対応付けによる $\boldsymbol{C}$ の推定精度低下を防ぐ．

また，提案システムは，ここで推定された $\boldsymbol{C}$ に基づき，特徴点マッチングと $\boldsymbol{C}$ の推定を再度行い $\boldsymbol{C}$ を更新することで，その精度をさらに向上させる． i 回目の対応点探索における探索範囲は，式 (8) をもとに次のように設定する．

$$\sqrt{|X_i - X_j|^2 + |Y_i - Y_j|^2} < \frac{R}{i} \tag{11}$$

以上のように，提案システムは，マーカーを用いた手法と特徴点マッチングによる手法を組み合わせることで変換行列 $\boldsymbol{C}$ の正確な推定を行う．これに基き，テンプレート画像中で選択されたネットワーク機器の部位が，映像中のどこに投影されるかを式 (4) から求めることで，ネットワーク情報の効果的な AR 可視化の実現を図っている．

4. 実験

本稿で提案した，映像中でのネットワーク機器の位置・姿勢の推定手法を実装し，その有効性を検証する実験を行った．

図 6 実験で用いたネットワーク機器の画像（1280 × 720 画素）

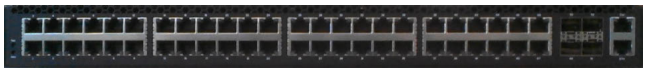


図 7 実験で用いたテンプレート画像（960 × 94 画素）

4.1 実験環境

実験には以下の環境を用いた．

- PC
  - CPU: Intel Core i7-4790 @ 3.60GHz
  - メモリ: 8GB
  - OS: Windows 7 Professional (x64)
  - 開発環境: Microsoft VisualStudio 2010
- カメラ
  VAIO Duo 11 のフロントカメラ（1280 × 720 画素）

提案手法での特徴点の抽出には，Scale Invariant Feature Transform (SIFT) [5] を用いた．各特徴点は 128 次元の特徴量で記述し，特徴点同士の類似度（相違度）を 128 次元の特徴量のユークリッド距離で評価した．決定された特徴点対から変換行列 $\boldsymbol{C}$ を求めるためには RANSAC アルゴリズム [7] を用いた．また，これらの実装には OpenCV[8] の C++ライブラリを利用した．

対象として用いたネットワーク機器をカメラで撮影した画像（1280 × 720 画素）を図 6 に，これを正面から撮影したテンプレート画像（960 × 94 画素）を図 7 に示す．両図は，ともに歪み補正を行った後の画像である．

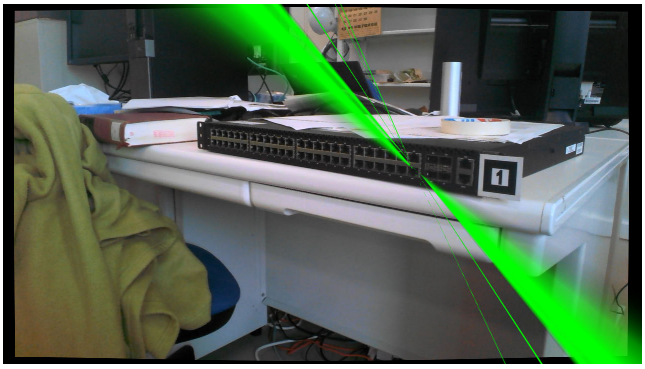
また，実験では，管理用ラベルが所定の箇所から最大 $\sqrt{t_x^2 + t_y^2} = 10\mathrm{mm}$，および $\theta = 10°$ ずれているものとした．

4.2 実験結果

図 6 の映像に対し，ネットワーク機器の位置・姿勢の推定を行った結果を図 8 に示す．図 8 では，推定結果に基づき，テンプレート画像中で得られた各ポートの位置を緑の線色で映像に重ね合わせ投影している．図 8 (a) は，管理用ラベルのみを利用し推定を行った結果である．この結果から分かるように，ラベルから離れた箇所（機器の左側）では，投影したポートの位置が実際の位置から大きくずれており，各ポートの状態を当該ポートの近傍に表示するような可視化を実現するためには位置・姿勢の推定精度が十

©2015 Information Processing Society of Japan

(a) 管理用ラベルのみを利用した推定

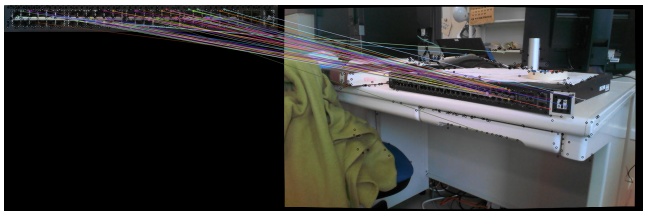


(b) 特徴点マッチングによる推定（探索範囲の制限なし）

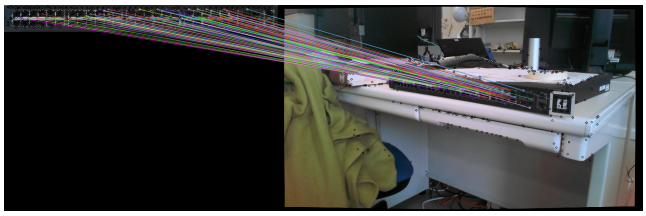
(c) 特徴点マッチングによる推定（探索範囲の制限あり）

(d) 提案手法（変換行列 $\boldsymbol{C}$ を 1 回更新）

図 8　ネットワーク機器の位置・姿勢の推定結果



(a) 探索範囲の制限無し



(b) 管理用ラベルを用いた推定の結果により探索範囲を制限

図 9　特徴点マッチングの結果

分ではないことを確認できる．図 8 (b) と (c) は，特徴点マッチングにより推定した結果である．図 8 (b) では，特徴点マッチングの際に探索範囲を制限せず，図 8 (c) では，管理用ラベルを用いた推定結果に基づき探索範囲を制限している．各々の特徴点マッチングの結果 50 組を図 9 (a) と (b) に示す．この結果から分かるように，探索範囲を制限しない場合，特徴点マッチングの誤りが多数生じるため，機器の位置・姿勢の推定精度も非常に低くなっている．これに対し，探索範囲を制限すると，特徴点マッチングの誤りを減らすことができ，機器の位置・姿勢も高い精度で推定できている．図 8 (d) は，提案手法による推定結果である．前回の推定結果に基づき，探索範囲を制限した特徴点マッチングを再度行い変換行列 $\boldsymbol{C}$ を更新することで，図 8 (c) に示した結果よりも，さらに高い精度で機器の位置・姿勢が推定されていることが確認でき，これは，推定精度向上に対する提案手法の有効性を示すものである．

次に，対象がカメラから離れている場合と，対象にネットワークケーブルが接続されている場合の映像について，提案手法により位置・姿勢を推定した結果を図 10 に示す．このような状況においても，各ポートの位置が映像中に正しく投影されていることが確認できる．

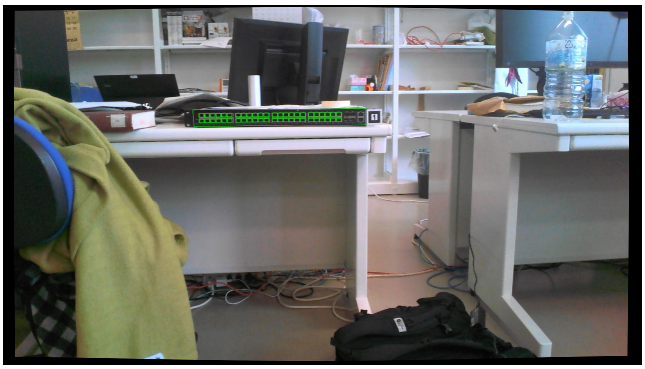
最後に，図 8 (d) の結果を得るために提案手法が各処理段階で要した処理時間を表 1 に示す．提案手法で今回用いた SIFT は，計算コストの高い手法であるため，SIFT による特徴点抽出と特徴量記述が，全体の処理時間の半分以上を占めている．

5.　おわりに

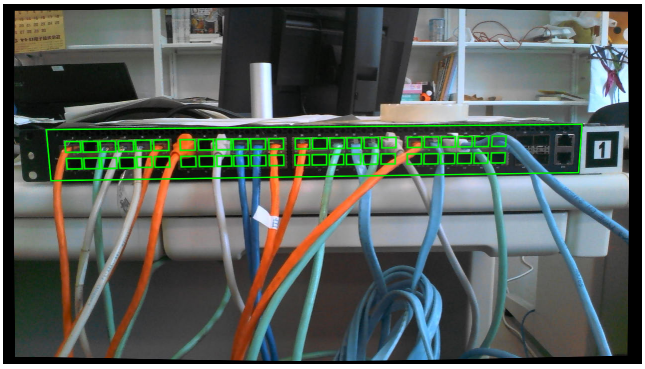
本稿では，ネットワーク情報の AR 可視化を行うためのシステムを提案し，これを実現する際に必要となる，映像中でのネットワーク機器の位置・姿勢推定手法を提案した．

実際のネットワーク機器の映像を対象とした実験から，提案する推定手法を用いることで，各ポートの状態を当該

©2015 Information Processing Society of Japan



(a)



(b)

図 10 (a) 対象がカメラから離れている場合の推定結果，(b) 対象にネットワークケーブルが接続されている場合の推定結果

表 1 処理時間

歪み補正	60ms
管理用ラベルを利用した位置・姿勢推定	37ms
SIFT による特徴点抽出（1452 個）	282ms
特徴量記述	291ms
特徴点マッチングによる位置・姿勢推定	118ms
変換行列 $\boldsymbol{C}$ の更新	105ms
その他	38ms
全体	931ms

ポートの近傍に表示する等，ネットワーク情報を適切な箇所に AR 可視化するために十分な位置・姿勢の推定精度が得られることが確認できた．

今後は，リアルタイムの AR 可視化を可能とするために，位置・姿勢推定処理の速度の向上を図る．そのための具体的な方法として，前フレームでの推定結果を利用した処理量の削減，マルチスレッド化による処理の高速化などが考えられる．また，AR 可視化をより効果的なものとするために，機器の位置・姿勢だけでなく機器に接続されたケーブル等を認識する手法の実現を進め，より多様なネットワーク機器設置環境において定量的観点から提案システムの有効性を検証する実験を行う予定である．

参考文献

[1] 和田佳 他, “実空間指向ネットワーク障害診断システムの実装と評価,” 第 72 回情報処理学会全国大会論文集, pp.274-274, 2010.

[2] 加藤博一, “拡張現実感システム構築ツール ARToolKit の開発,” 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解 101(652), 79-86, 2002-02-14

[3] 廣瀬直子 他, “詳細な形状モデルを用いたオブジェクトの 3 次元位置・姿勢推定,” DEWS2006 in Ginowan, IEICE Data Engineering Technical Group, 4B-i10, 2006

[4] l. Vacchetti, ”Combining Edge and Texture Information for Real-Time Accurate 3D Camera Tracking,” ISMAR ’04 Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, 2004

[5] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int.Journal of Computer Vision, Vol.60, No.2, pp.91-110, 2004

[6] 高木雅成 他, ”SIFT 特徴量を用いた交通道路標識認識,” 電気学会論文誌, Vol.129-C, No.5, pp.824-831, 2009.

[7] M. A. Fischler and R. C. Bolles, ”Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” Commun. ACM,vol.24, no.6, pp.381-395, June 1981.

[8] OpenCV, http://opencv.org/

©2015 Information Processing Society of Japan