

# Support Vector Machine による 物体認識を利用した保守訓練支援の試み

中島 慶人

## 1. はじめに

バーチャルリアリティ (VR: Virtual Reality) を電力業務へ導入すると、安全に各種の保守訓練が行えると期待され、VR 技術の活用方法が模索されている。しかし、全てをコンピュータ・グラフィックス (CG: Computer Graphics) で作成する VR は、設備の操作感覚や現場の臨場感に欠けるため、電力会社に限ると未だ広報業務の利用に留まっている。

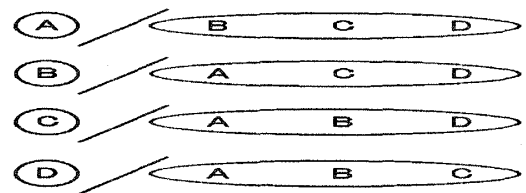
そこで、我々は VR 技術普及の障害の 1 つである操作感覚欠如を実設備の一部に CG を合成することで補い、保守や訓練業務の効率化に向けて VR の活用を検討している [1, 2]。この研究の中で、カメラで撮影した設備を認識し CG の合成位置を特定する画像処理が必要となった。しかし、従来の VR システムのように認識設備ごとに特殊な認識プログラムを作成したり、物理的なセンサや特殊マーカを準備したりすると、事前準備に費用と手間がかかってしまう。そこで、機械学習を利用し、事例を学習させるだけで認識対象を簡単に追加・更新できる方式を検討した。その中で、機械学習の一方式であるサポートベクターマシン (SVM: Support Vector Machine) を利用したので、その概略を紹介する。

## 2. SVM による多クラス分類

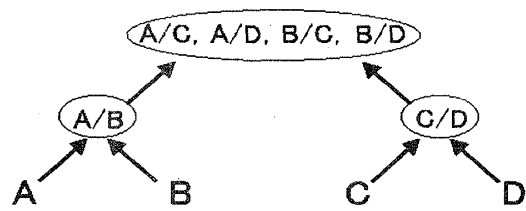
SVM は 2 クラス分類問題用の学習方式であるため、複数の対象を分類するには SVM を組合せて利用する必要がある。そのような組合せ方式として、図 1 に示す方式が提案されている [3~5]。図 1 は 4 クラス (A, B, C, D) の分類方式を示し、本論では図 1-a, b, c の各方式を図の書き方から便宜上、一対他方式 (One vs. All others), ボトムアップ方式, トップダウン方式と呼ぶことにする。また、本章では分類したいクラ

スの数を  $n$  とする。

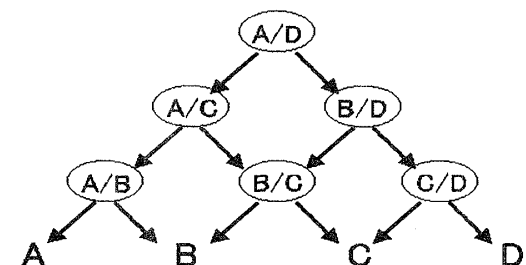
一対他方式 (図 1-a) は、1 つのクラスに対し残り全てのクラス ( $n-1$ ) を 1 クラスと見立て、SVM によるトレーニングを行う方式である。例えば、A と残り全て (BCD) の間で SVM のトレーニングを実行し、超平面を決定する。B, C, D に対しても同様の処理を行う。そのため、 $n$  クラスに対し  $n$  回のトレーニングを行う必要がある。また、新規データの分類も、各トレーニング結果と新規データを比較するため、SVM による  $n$  回の判定が必要となる。ただし、 $n$  回の判定の結果、複数のクラスが候補として残った場合、あるいはどのクラスにも判定されなかった場合、超平面からの距離により最終的なクラスを選ぶ必要がある。



a. 一対他による比較方式



b. ボトムアップによる比較方式



c. トップダウンによる比較方式

図 1 SVM による多クラス分類方式

なかじま ちかひと

財電力中央研究所

〒 201-8511 狛江市岩戸北 2-11-1

ボトムアップ方式 (図 1-b) は、テニストーナメントのような勝ち抜き選抜方式であり、SVM によるトレーニングを 2 クラス間で行う方式である。この方式は、あらかじめ全てのペアに対しトレーニングを行う必要があるため、 $n(n-1)/2$  回のトレーニングが必要となる。新規データの判定も 2 クラス間で行われ、最終結果を得るまでに  $n-1$  回の判定が行われる。

トップダウン方式 (図 1-c) は、ボトムアップ方式と同様にあらかじめ全てのペアに対しトレーニングを行う必要がある。新規データの判定はグラフの最上位ノード (A/D, クラス A とクラス D の判定) から分類が実行され、判定されたクラスの方へ向かい処理が進められる。例えば、最上位で A が選ばれた場合、次の層では A 側のノード (A/C) が判定される。同処理を繰り返し最後に到達したクラスが認識結果となる。そのため、新規データの判定回数は  $n-1$  回となる。

### 3. 服装による人物認識

保守訓練用のシステムでは、熟練度に応じて訓練者ごとに異なるシナリオを準備する予定である。このシナリオを訓練者ごとに自動的に切り替えるためには、訓練者の認識が必要となる。すでに、SVM を使い顔の検出や認識が行われている [6, 7] ため、それらを将来的に導入する予定で、まずは服装を利用した人物認識を検討した [8]。

#### 3.1 認識に利用する特徴

SVM による人物検出方式が提案されている [9] が、SVM 単独で画像内の人物を認識しその位置を検出するには、画像内を SVM により探索しなければならず処理に時間がかかる。そのため、遅延時間の短い応答を必要とする VR システムにはそれらの方法を直接活用できない。

そこで、SVM による人物検出を行わず、画像の背景差分とフレーム間差分処理により人物領域を抽出した後、人物位置に影響を受けない色情報を使い、画像内の人物を SVM により認識することにした。なお、この SVM による人物認識では、下記の 4 種類の色特徴により認識率を比較した。

##### (1) RGB

人物領域の RGB 値を色ごとにヒストグラム化する。各色を 32 分割し、合計 96 次元のヒストグラム特徴を画像から算出する。

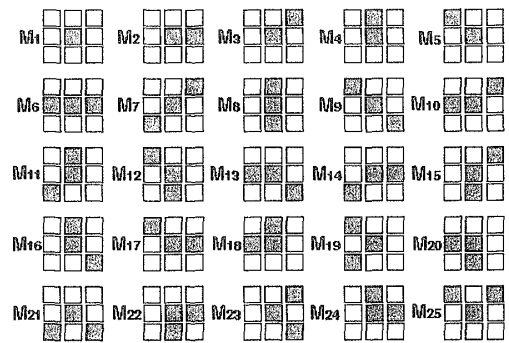


図 2 局所自己相関の局所パターン [10]

##### (2) 正規化 rg

RGB の値から  $r = R/(R+B+G)$ ,  $g = G/(R+B+G)$  を計算し、この 2 次元の rg に対応するヒストグラムを求める。2 次元の r と g の座標をそれぞれ 32 分割し、合計 1024 次元 ( $32 \times 32$ ) の特徴を画像から算出する。

##### (3) RGB+人物のサイズ

人物の身長や横幅なども人物認識に使えるのではないかと考え、(1) の RGB のヒストグラムに、人物領域の縦横の画素ヒストグラムを加えた。人物を抽出した画像に対し、縦 10 分割、横 30 分割し、縦横独立に画素数のヒストグラムを求め、RGB と合わせ合計 136 次元 ( $32 \times 3 + 10 + 30$ ) の特徴を算出する。

##### (4) 局所自己相関

人物領域の色の局所自己相関を特徴として計算する。この局所自己相関の計算には、図 2 に示す局所パターン [10] を利用する。ただし、脳研究の最近の成果に、視覚野の方向を感知する方向受容野群の働きが MAX オペレーションで説明できるとの仮説 [11] が提案されているため、2 次の相関 (図 2 でのパターン  $M_6 \sim M_{25}$ ) を方向受容野と見立て、最大値を示す相関だけを特徴に加算した。本実験では画像から合計 75 次元 ( $25 \times 3$ : 25 パターンに 3 種類の色) の特徴を抽出する。

#### 3.2 人物と向きとの認識

SVM による 3 種類の多クラス分類方式と 4 種類の特徴を使い人物と向きとの認識評価実験を行った。

実験では 4 人の被験者から、前後左右のトレーニング画像を各 40 枚 (合計 640 画像) 撮影し、さらに、その各方向の画像から 10 枚ずつを選び、被験者ごとに 40 画像を人物トレーニング用に準備した。それらの画像により人物と個人ごとの向きを別々にトレーニングした。

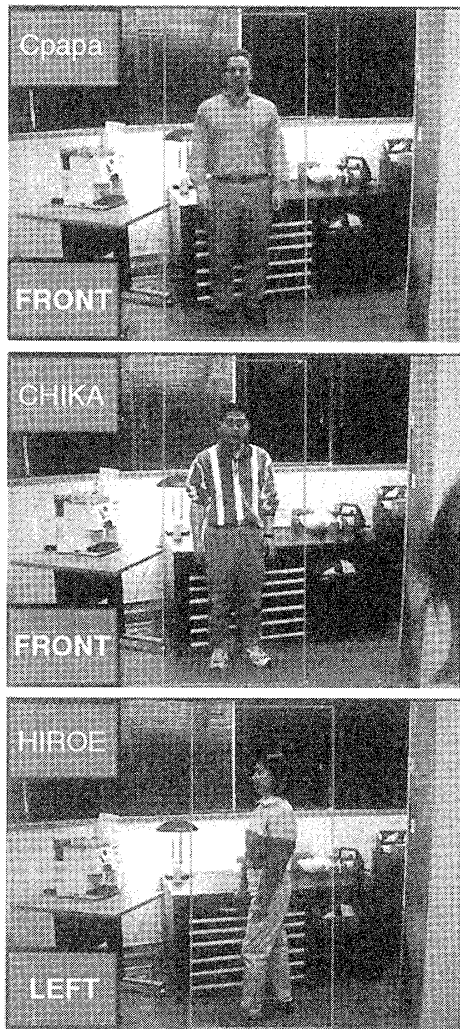


図3 服装による人物と向き認識の実験結果

新たに撮影した4人合計418画像にSVMにより人物と向きの認識を行った。図3に実験結果例を示す。図3の各画面の左上ウインドウが人物の認識結果(名前)であり、左下ウインドウが向きの認識結果である。また、人物を囲む長方形の枠はシステムが捉らえた人物位置である。システムは初めに人物を認識し、次に向きを判定する段階的な認識を行っている。図4に各方式と各特徴による人物の認識結果を、図5に向きの認識結果を示す。なお、他方式との比較として、図4と図5にはK-最近隣法(K-NN: K-Nearest Neighbor)による認識結果を示した。なお、この実験で使用したSVMはいずれも線形カーネルである。

### 3.3 16日間の人物の認識

図4と図5で正規化rgが最も高い認識結果を示したため、今度は16日間にわたり向きを考慮せず人物を自動的に撮影した画像1127枚数(8人分)を準備し、SVMによる認識実験を行った。得られた画像を

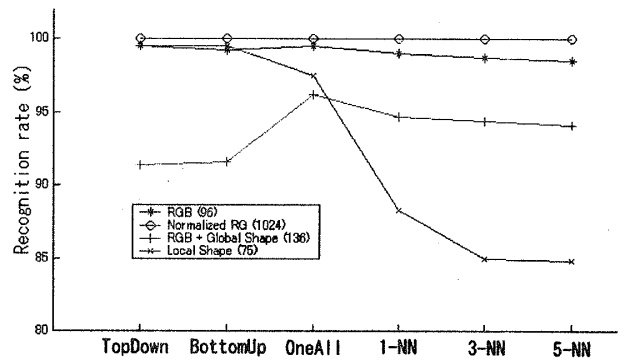


図4 多クラス分類方式と4種類の特徴を使った人物認識の結果比較

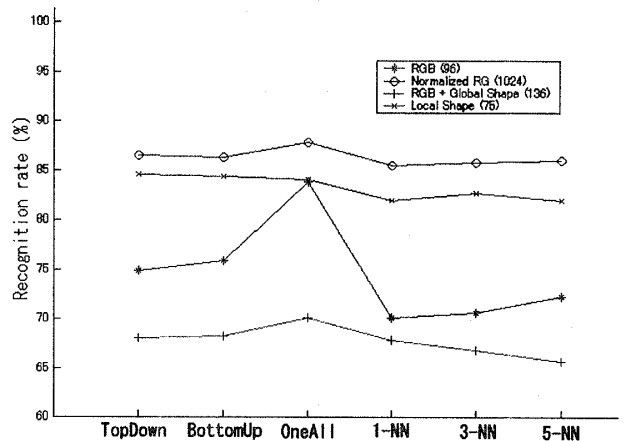


図5 多クラス分類方式と4種類の特徴を使った向き認識の結果比較

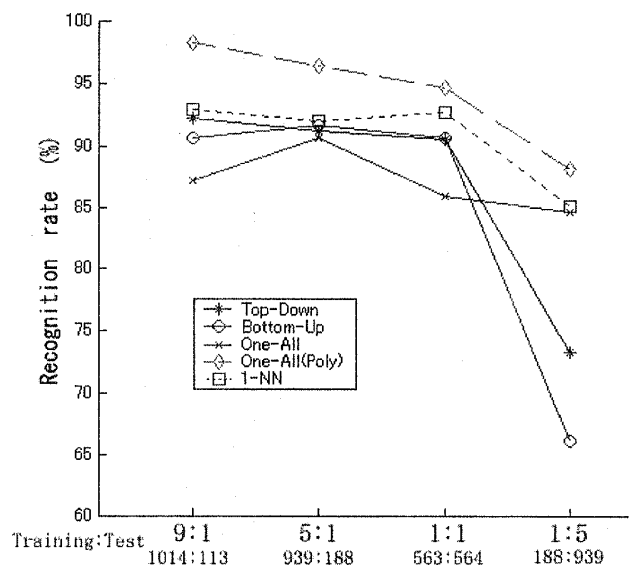


図6 多クラス分類方式と4種類の特徴を使った人物認識の結果比較

トレーニング用データと認識実験用データに、9対1, 5対1, 1対1, 1対5に分け、線形のSVMと多項式カーネルのSVMで認識実験をした結果が図6である。

## 4. 物体認識

本章では人物認識と同様に、物体認識に同システムを適用した結果を示す。ただし、実験での多クラスにはSVMによるトップダウン方式を利用し、特徴には正規化rgを利用した結果である。

### 4.1 物体認識

この実験では、あらかじめ5つの物体(図7のような開閉弁か碍子など)を手にとり各100枚(合計500枚)のトレーニング画像を撮影した。この画像を使い、各クラス間でSVMのトレーニングを行った。

トレーニング画像とは別に、新たに各400枚(5物体で合計2000枚)の画像を撮影し、物体認識実験を行った。図7に試作システムの実験結果例を示す。図7での各画像の左下ウィンドウはシステムが認識した物体名を示している。なお、左上ウィンドウは物体の検出位置[12](本論では省略する)を表示している。



図7 設備認識の実験結果例

SVMによる5物体の認識率は100%であった。

### 4.2 物体認識特性

システムの物体認識特性を調べるため、コロンビア大学の画像ライブラリCOIL (Columbia Object Database Library[13])の7200画像を使い認識実験を行った。COILは、物体を5度ごとに回転させ撮影した100物体(合計7200枚=100×360°/5°)の画像が収録されている実験用画像ライブラリである。図8に画像例を示す。

各物体の画像数の5%、10%、20%、30%、40%をSVMで学習を行い、残りの画像で認識実験を行った。その結果を、K-NNと2次の多項式カーネルによるSVMの認識結果とともに図9に示す。なお、K-NNはK=1, 3, 5を実験したが、K=1の結果がK=3, 5を常に上回ったため図9にはK=1の結果のみを記載した。図9に示すように、1-NNよりもSVMの認識率は常に上であった。また、学習データが少ない場合、多項式カーネルを用いたSVMよりも線形のSVMの方が良い結果を示すが、学習データが増えるに従い結果が入れ替わる。これは、ある程度サンプル

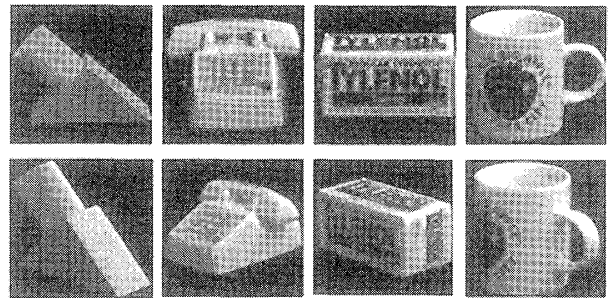


図8 データベース COIL の画像例

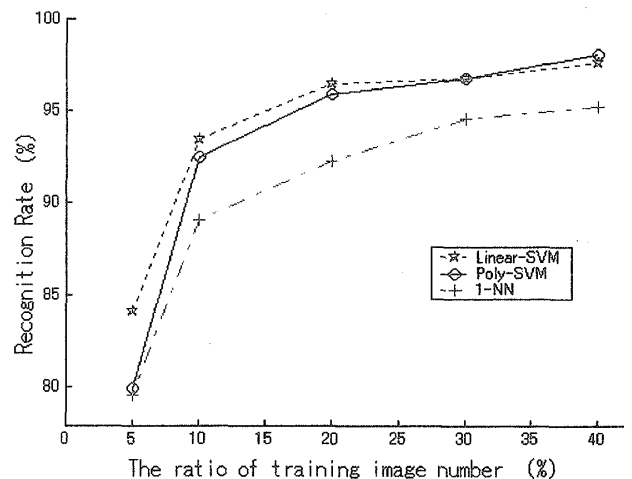


図9 データベース COIL の100物体の認識結果

数が確保できる場合、高次空間にデータを写像し分類する方式が有効であることを示している。

学習データ 5% (400 枚) の認識実験で、残り 6800 枚の認識時間は SGI-02 (250 MHz) のコンピュータを使い、線形の SVM が 6 分 50 秒、1-NN は 35 分 9 秒であった。

## 5. 設備画像と解説の合成

4 章での実験と同様にトップダウン方式の SVM と正規化 rg を使い設備を認識した後、認識結果に応じて解説内容を更新している様子を図 10~12 に示す。いずれも、稼動システムの画面コピーである。

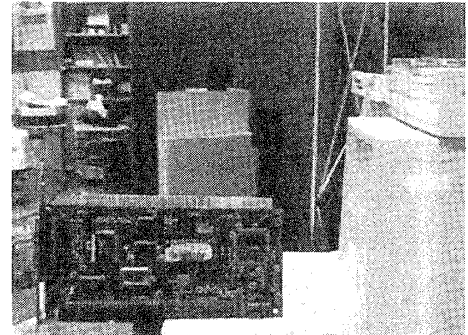
図 10 は開閉弁が利用されている状況を、図 11 は基板のセットされるべき場所を示している。また、図 12 は、見る方向に応じてシステムが説明内容を自動的に変更している様子を示す。なお、開発したシステムは説明画像を眼鏡型ディスプレイに表示するだけでなく、音声による説明機能も含んでいる。

## 6. おわりに

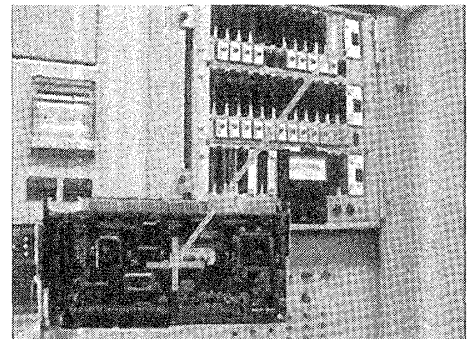
本論では現在研究を進めている VR による保守訓練支援システムの一部として作成した、SVM による複数の物体認識方式を示した。この方式は、認識対象

ごとに認識プログラムを作成したり、特殊なセンサを使ったりすることなく、システムに与える学習事例を変えることで物体認識を行えるものである。

我々の認識システムは、物体認識に色情報のみを用いるため、色が似ている場合や照明が変化した場合、認識率が悪くなる。特に、同じような作業服を着てい

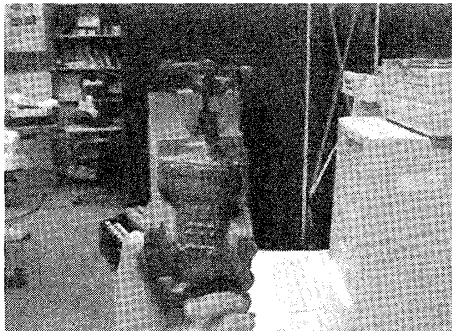


a. カメラの撮影画像

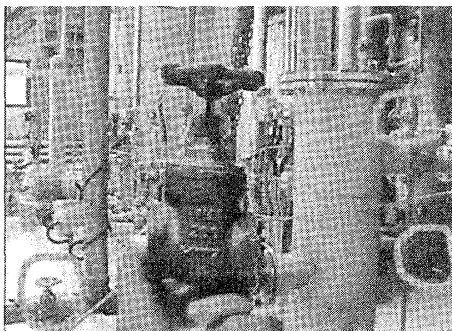


b. 基板の挿入位置を示す説明用画像との合成

図 11 基板を認識し説明を表示している例

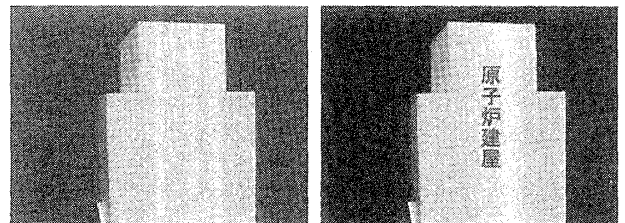


a. カメラの撮影画像



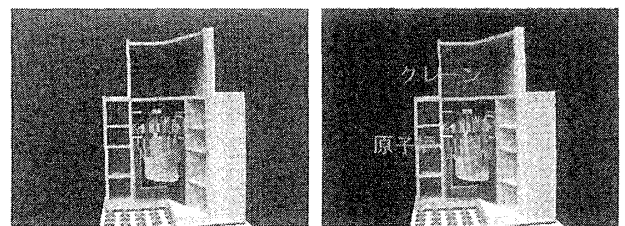
b. 開閉弁が使われている場所を示す画像との合成

図 10 開閉弁を認識し説明を表示している例



a. カメラの撮影画像

b. 認識に基づく解説



c. カメラの撮影画像

d. 認識に基づく解説

図 12 向き認識により解説内容を自動更新した様子

る場合、服による人物認識ができない問題がある。今後、顔領域の特徴や色以外の形状特徴も利用し認識精度向上を図る予定である。また、本方式は入力物体を既知の物体に割り振る方式のため、未知の物体を自動的に学習する機構を持たない。今後、未知物体の自動学習機構を画像検索技術との組合せで検討する予定である。

#### 参考文献

- [1] 渡部直人, 萩原豊, 中島慶人, 伊藤憲彦: 「バーチャルリアリティ技術による放射性物質輸送総合支援システムの概念検討」, 電力中央研究所研究調査報告, No. U 99052 (2000)
- [2] 伊藤憲彦: 「複合現実のための画像認識技術, 位相限定相関を用いた3次元物体の位置・姿勢推定方法」, 電力中央研究所研究報告, No. R 98011 (1999)
- [3] Chapelle O., Haffner P. and Vapnik V.: “Support Vector Machines for histogram based image classification”, IEEE Trans. Neural Networks (1999)
- [4] Platt J., Cristianini N., and Shawe-Taylor J.: “Large Margin DAGs for Multiclass Classification”, Advances in Neural Information Processing Systems (2000)
- [5] Pontil M. and Verri A.: “Support Vector Machines for 3-D Object Recognition”, IEEE Trans. PAMI (1998)
- [6] Heisele B., Poggio T. and Pontil M.: “Face detection in still gray images”, MIT A. I. Memo, No. 1687, C. B. C. L. No. 187 (2000)
- [7] Kumar, V. and Poggio T.: “Learning-based Approach to Real Time Tracking and Analysis of Faces”, Proceedings of the Fourth International Conference on Face and Gesture Recognition, pp. 91-96 (2000)
- [8] Nakajima C., Pontil M., Heisele B. and Poggio T.: “People Recognition in Image Sequences by Supervised Learning”, MIT A. I. Memo, No. 1688, C. B. C. L. No. 188 (2000)
- [9] Papageorgiou C. and Poggio T.: “Trainable Pedestrian Detection”, Proc. of ICIP (1999)
- [10] 大津展之, 栗田多喜夫, 関田巖: 「パターン認識」, 朝倉書店 (1996)
- [11] Riesenhuber M. and Poggio T.: “Hierarchical models of object recognition in cortex”, Nature Neuroscience, 2(11), pp. 1019-1037 (1999)
- [12] Nakajima C., Itoh N., Pontil M. and Poggio T.: “Object recognition and detection by a combination of Support Vector Machine and Rotation Invariant Phase Only Correlation”, Proc. ICPR (2000)
- [13] Nene S. A., Nayar S. K., and Murase H.: “Columbia Object Image Library (COIL-100)”, Technical Report CUCS-006-96 (1996)