Vol.2019-CVIM-215 No.23
2019/1/18

ライトフィールドレンダリングを用いた距離変化にロバストなLFD特徴量

志礼田 賢人[1,a)]　Xu Yichao[2]　長原　一[2]　谷口　倫一郎[1]

概要：画像をコンピューターに認識させる物体認識のタスクにおいて，認識が困難な対象として透明物体が挙げられる．透明物体は背景を透過し，シーンによってその外観が全く異なるためである．この問題について，Xu らは透明物体によって生じる歪みを特徴とする LFD 特徴を用いて，シーンに依存せずに認識を行う手法を提案した．しかしながら，LFD 特徴は物体とカメラの位置関係が異なると認識率が大きく低下するという問題があった．本研究ではライトフィールドレンダリングを用いて距離の違いによる LFD 特徴の変化を補正することにより，より安定して透明物体認識を行う手法を提案する．

キーワード：透明物体，ライトフィールド，物体認識，画像特徴量

LFD feature robust to distance change using light field rendering

Kento Shireida[1,a)]　Xu Yichao[2]　Hajime Nagahara[2]　Rin-ichiro Taniguchi[1]

1. はじめに

画像をコンピューターに認識させる物体認識は盛んに研究が行われている分野であり，ロボットによる物体の認識やハンドリングなど，多くのアプリケーションの根幹をなす要素として，今後ますます重要性が増してゆくと考えられる．物体認識の内，画像に正しいラベルを付与する画像分類の手法としては，画像から計算された特徴量をベースにした手法や，Deep Learning を用いた手法などが提案されている．

画像分類によく用いられる特徴量としては，Haar-like[1] や HOG[2]，Edgelet[3] 等が挙げられる．Haar-like 特徴量は 2 つの領域の平均輝度の差によって定義され，これは局所的なエッジ・線成分を表現する．差分を用いているため，画素値をそのまま用いる場合に比べて照明変動等の影響を受けにくいという利点があり，顔や物体の検出などに使用されている [4][5]．HOG 特徴量は局所領域の輝度値の勾配方向をヒストグラムとしたものである．これら局所特徴量は，輝度勾配を用いていることで画像サイズの変化や照明変動に対してロバストだというメリットを持ち，人物検出などに用いられる [6]．Edgelet 特徴量は局所領域におけるエッジのつながりにより定義され，これは局所領域内の特定の形状を表現する．これにより正確な形状を捉えることが出来るという利点があり，人物追跡などに活用されている [7]．Deep Learning を用いた手法としては，AlexNet[8] や ResNet[9] 等が高い認識率を実現した．

これらの手法は物体のテクスチャに基づいて認識を行っているが，テクスチャベースの手法が適用できない対象として透明物体が挙げられる．これらの手法は，物体の外観はシーンによって大きく変化しないことを前提としているが，透明物体は背景を透過するため，背景が変化するとその外観も変化する (図 1)．そのため，背景に依存しない物体固有の特徴を元に認識を行う必要がある．

そこで本研究ではライトフィールドを用いることで背景に依存しない物体固有の特徴を取得する．ライトフィールドとは光線によって構成される場のことである．カメラによってシーンを撮影する際，レンズに入射する光線は通過位置 (s,t) と角度 (u,v) の 4 次元のパラメータで表現する

[1] 九州大学大学院システム情報科学研究院
[2] 大阪大学データビリティフロンティア機構
[a)] shireida@limu.ait.kyushu-u.ac.jp

図 1 異なるシーンにおける透明物体．背景が異なると透明物体の外観も異なる．

ことが出来る．通常のカメラは光線の通過位置を複数記録することが出来ないため位置の情報は失われてしまうが，ライトフィールドカメラという特殊なカメラを用いることでこの位置情報を記録することが出来る．この光線空間の情報を用いることで，撮影後のフォーカス位置の変更や，複数視点の画像の生成などのアプリケーションが実現できる．

Xu ら [10] はライトフィールドを取得して透明物体の分類を行った．この手法ではライトフィールドカメラによって撮影した複数視点画像を元に，透明物体によって生じる歪みを求め，この歪みを特徴量としている．この手法は高い認識率を実現しているが，撮影時のカメラと透明物体の距離が訓練データと一致する物体しか高精度で認識できない．そこで本研究はライトフィールドカメラで撮影されたデータを元に，ライトフィールドレンダリングにより認識対象の画像と一致するような仮想視点画像を生成してそれを用いることで，距離の変化による認識率の低下を抑制する手法を提案する．

2. Xu らの手法

Xu ら [10] は透明物体によって生じる歪みを利用して透明物体の分類を行った．背景とカメラの間に透明物体が存在する場合，透明物体の部分において背景が歪められた画像が得られるが，この歪みは背景のテクスチャに依存せず，物体の形状・材質によって決まる量であり，物体の特徴を表現していると考えられる．このことから，Xu らはこの透明物体によって生じる歪みを特徴量として透明物体の認識を行った．

歪みの算出にはライトフィールドカメラから得られた複数視点の画像を用いている (図 2)．格子状に配置された複数視点の内，中心の視点の画像を基準として，他の視点との間のオプティカルフローを計算する．視点が変化すると視差が生じて物体や背景が画像中の異なる位置に写るため，この移動量がフローとして求まる．

この時，背景部分の画素のオプティカルフローは視点の変化に対して線形な値を持つのに対して，透明物体部分の

図 2 ライトフィールドカメラ (左) と取得される複数視点画像 (右)．

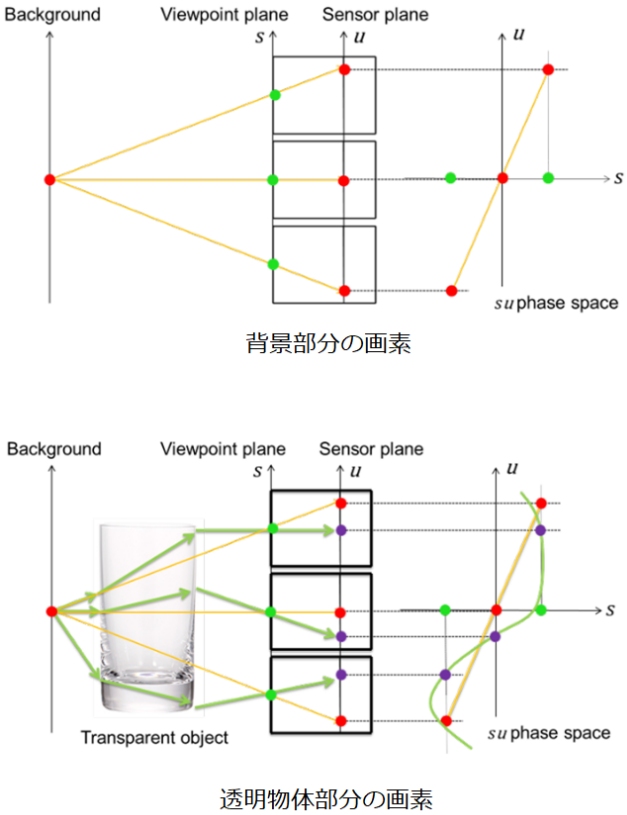

図 3 オプティカルフローの線形性．

画素は透明物体による歪みの影響を受けるため非線形な値を持つ (図 3)．この非線形なオプティカルフローは透明物体の特徴を表現した値であると考えられるため，Light Field Distortion(以下，LFD) と呼び，認識に利用している．画素 (u,v) における LFD は，視点位置を表す変数 (s,t) とオプティカルフロー $(\Delta u, \Delta v)$ で以下のように表される．

$$LFD(u,v) = \{(s,t,\Delta u,\Delta v)|(s,t) \neq (0,0)\}$$

Xu らの手法は，訓練に用いた画像と認識対象の画像が同一のセッティングで撮影された場合には高い分類精度を実現した．しかし，認識対象の画像が撮影された時のカメラと物体間の距離が，訓練画像における距離から離れるほど，分類精度が大きく低下してしまうという問題点があった (図 4)．これは，カメラと物体間の距離が変化することで視差やオプティカルフローの値が変化し，全く異なる LFD が得られるためである．そこで本研究ではこの認識

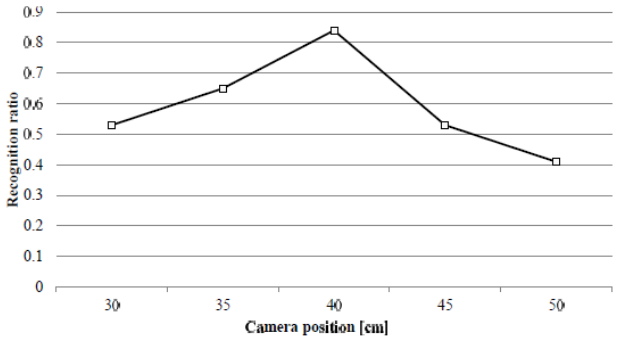

図 4 カメラと物体間の距離に対する認識率の推移.

率の低下を抑制するため，撮影したライトフィールドを利用して仮想視点画像を作成し，距離の変化に対応する手法を提案する．

3. 提案手法

本手法では取得したライトフィールドを元に，Light Field Rendering を用いて学習時とは異なる視点の画像を仮想的に生成する．これにより，認識対象の画像と同一の視点の画像を生成することで，距離の変化による LFD の変化を補償する．そして生成した仮想視点画像から得られた LFD 特徴を学習時のパターンに変換しマッチングに用いることで，認識率の低下を抑制する．認識手法のアルゴリズムは大きく 5 行程に分けられ，(1) 仮想視点画像の生成，(2) 特徴量の生成，(3) マスキング，(4) 量子化，(5) マッチング，の順に行われる．認識は典型的な Bag of feature の手法に従って行う．このアルゴリズムの全体像は図 7 のようになる．

3.1 仮想視点画像の生成

まず取得したライトフィールドを元に仮想視点画像を生成する．仮想視点画像の生成は，Light Field Rendering[11] によって行う．取得したライトフィールド LF は光線の通過位置 (s,t) と角度を (u,v) を変数に持つ関数として $LF(s,t,u,v)$ と表せる．ここで中心視点を $LF(0,0,u,v)$ と表すこととし，この視点を光軸上で物体に d だけ近い位置に平行移動することを考える．この仮想的な視点において記録されるライトフィールド $LF(s',t',u',v')$ は以下の式で表せる．

$$LF(s',t',u',v') = LF(-d*tan(u), -d*tan(v), u, v)$$

従って仮想視点において記録される光線を記録済みの LF から求めることで，仮想的な視点の画像を生成することが出来る (図 5)．ただし一般に (s',t') は整数にならず，対応する光線は存在しないため，取得された LF を元に 4 次元補間を行い近似的な値を求める．これにより認識対象の画像と同一の視点の画像を生成することで，認識対象の画像と同様の LFD を取得することが可能となる．

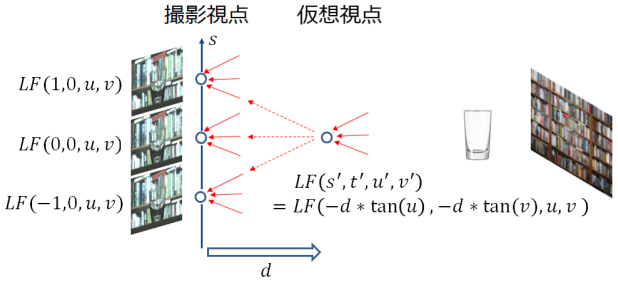

図 5 Light Field Rendering.

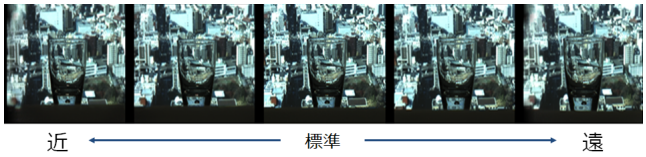

図 6 標準視点画像（中央列）と仮想視点画像（他列）.

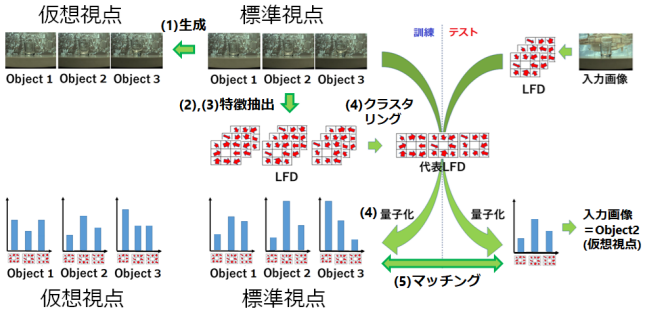

図 7 提案手法のアルゴリズム.

Light Field Rendering により作成された仮想視点画像の例を図 6 に示す．三列目が標準の視点画像であり，左に行くほど物体に近い仮想視点，右に行くほど物体から遠い仮想視点画像である．

3.2 特徴量の生成

複数視点の画像において，視点が異なると視差が生じるため画像中の異なる位置に写る．ここで中心視点の画像の各画素が，他の視点の画像においてどの程度移動しているかを求める．推定には Brox らの Large-Displacement-Optical-Flow[12] を用いる．この手法は変分法と記述子マッチングを組み合わせた手法である．この手法を用いることで，中心視点の画像中の各画素毎に (視点数-1) 個のオプティカルフローが求まる．ここで，視点位置とオプティカルフローの組を，歪み特徴量 LFD として定義する (図 8)．画素 (u,v) における LFD は，視点位置を表す変数 (s,t) とオプティカルフロー $(\Delta u, \Delta v)$ で以下のように表される．

$$LFD(u,v) = \{(s,t,\Delta u,\Delta v) | (s,t) \neq (0,0)\}$$

この LFD を分類のための特徴量として用いる．

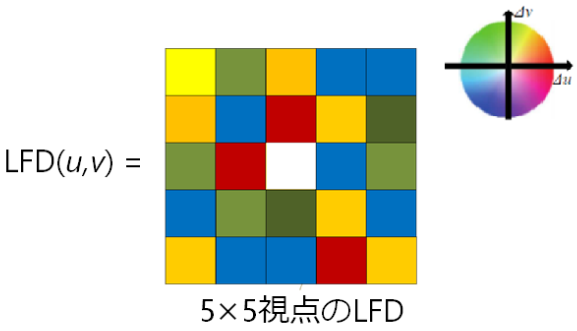

図 8 LFD(5 × 5)．色がオプティカルフローの向きと大きさを表す．

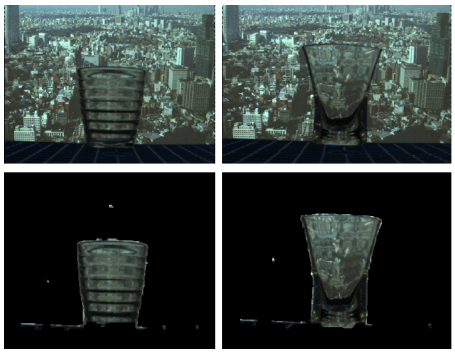

図 9 マスク例. 透明物体と判定された部分を白，背景を黒で示している．

3.3 マスキング

オプティカルフローは画像中の全画素について計算されるが，背景上の点を取り除き透明物体部分のみの LFD を用いて分類を行うことが望ましい．そこで，背景の点と透明物体の点のオプティカルフローの間には図 3 に示したような線形性の違いがあることを利用し，線形性を基準にしたマスキングを行い透明物体上の点のみを選別する．マスキングには TransCut[13] を用いる. この手法は，背景・物体の 2 値ラベル付け問題を解くことによって画像を画素ごとにセグメンテーションし，背景上の点を取り除いている (図 9)．そのためにデータ項と平滑化項を持つエネルギー関数を次式で定義し，これをグラフカットによって最小化する．

$$E(l) = \sum_{p} R_p(l_p) + \alpha \sum_{(p,q)} B_{p,q} * \delta(l_p, l_q)$$

α は平滑化項に対する重み，R がデータ項，B が平滑化項である．データ項と平滑化項を構成する主要な 2 つの要素は，LFD の線形性を表す e と，オクルージョンの確からしさ O である．

LFD は背景上の点については線形となり，透明物体上の点については非線形となる．1 つのピクセルにつき (視点数-1) 個の 2 次元のオプティカルフローが求まるため，各ピクセルについて 2 × (視点数-1) 次元の LFD が求まる．ここで，LFD の線形性を超平面のフィッティング誤差によって評価する．背景上の点のオプティカルフローは視点の変化に対して線形となるため，LFD は 4 次元空間上のある超平面上に分布し，フィッティング誤差は小さくなる．逆に透明物体上の点であれば，オプティカルフローは視点の変化に対して線形とならず，フィッティング誤差は大きくなる．従ってフィッティング誤差 e は，LFD の線形性を表現しているといえる．

また物体と背景の境界部分を判別し，認識に利用する．物体と背景の境界部分では，視点が変化するとオクルージョンが発生する．これは視点が変化した際に元の視点では見えていた領域が物体によって遮られるためである．このオクルージョンをオプティカルフローを利用して発見する．オクルージョンが発生した場合，視点が変化するとオプティカルフローの追跡に失敗するため，双方向でオプティカルフローを計算すると元のピクセルとは異なるピクセルにたどり着くこととなる．この双方向フローの誤差の大きさによって，そのピクセルにおいてオクルージョンが発生しているかを判定する．この時求まるオクルージョンの確からしさを O とする．

背景上の点では，e は小さく，O は小さくなる．境界部分では e と O ともに大きい．物体上では e は大きく，O は小さくなる．この関係から，データ項 R と平滑化項 B を次式のように定義し，ラベル付けを行う．

$$R(0) = \beta\tilde{e} * (1 - \tilde{O})$$

$$R(1) = \tilde{e} * \tilde{O} + (1 - \tilde{e})$$

$$B_{p,q} = exp(-\gamma(w_{p,q} + w_{q,p}))$$

R(0) はその画素を背景とするコスト，R(1) は物体とするコスト，$\tilde{e}$ と $\tilde{O}$ はそれぞれ正規化された e と O，w は隣接する画素 p, q の間でラベルが異なる際の重みを表す．β, γ は各項の重みである．これにより，境界部分と背景部分の画素には背景ラベル，物体部分の画素には物体ラベルが割り当てられる．

3.4 量子化

マスキングにより選別した LFD を用いて各物体を表現する．まず訓練画像から求まった LFD を，K-means クラスタリングによって K 個のクラスタに分類する．そして K 個のクラスタの重心となっている LFD を代表 LFD として決定する．この代表 LFD を基底として，各物体の画像から得られた LFD をそれぞれ量子化することで，代表 LFD によるヒストグラムを物体ごとに求める．これにより，各物体の画像中に代表 LFD がどのような割合で含まれるかを表現したヒストグラムを作成する．認識対象が写った入力画像についても，同一の代表 LFD を用いて量子化し，ヒストグラムを作成する．

図 10 ライトフィールドカメラ Lytro ILLUM.

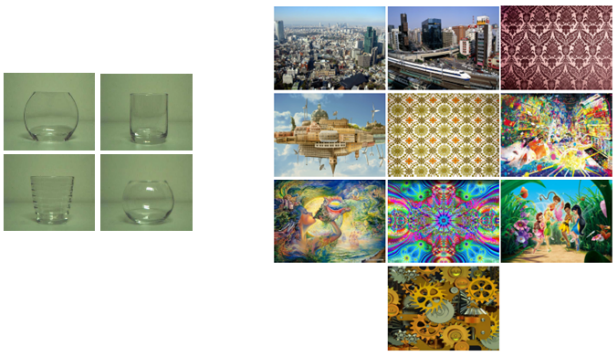

図 11 認識対象の物体と使用した背景画像.

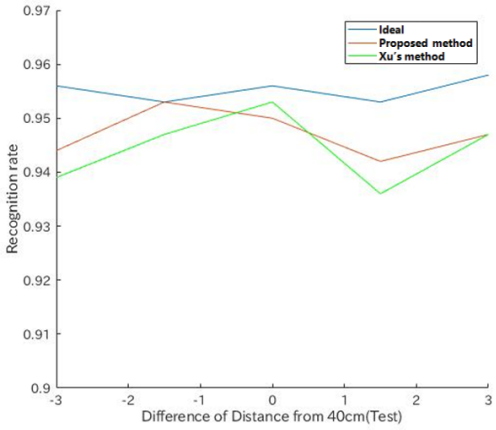

図 12 認識率の比較.

3.5 マッチング

訓練画像から得られた各透明物体のヒストグラムと，識別対象の物体から得られたヒストグラムとを比較し，最も類似するものを入力画像に写っている透明物体であると判定する．ヒストグラムの類似度は交差によって計算する．

4. 実験結果

今回用いている撮影デバイスのライトフィールドカメラ Lytro Illum(図 10) は，メインレンズとセンサの間にマイクロレンズを配置したカメラであり，一般のカメラにおいて積分されてしまう光線を独立に記録することが出来る．この情報を活用することで，撮影後のフォーカス位置の変更や複数視点画像の生成などが実現できる．

今回の実験では，視点数は縦 5，横 5 の 25 視点とする．LytroIllum で撮影されたデータを処理することで縦 13，横 13 の計 169 視点の画像が得られ，その中心の 25 視点を用いる．ここでは，図 11 に示した透明物体 4 種類と背景 10 種類を用いるという条件下で，Xu の手法による分類精度と提案手法の分類精度とを比較した．実験設定としては，カメラと物体との距離が 40cm，物体と背景との距離が 150cm となるように配置した状態を標準のセッティングとし，それに加えてカメラを物体に 1.5cm 刻みで近づけた 2 パターンと，遠ざけた 2 パターンの計 5 パターンのデータを取得した．これらの 5 パターンの画像を認識対象として，訓練には標準距離の画像と，生成された 4 パターンの仮想視点画像を用いた．背景としてはプロジェクタで画像を投影したスクリーンを用いた．

分類精度を比較した結果を図 12 に示す．横軸は認識対象の画像におけるカメラと物体間の距離，縦軸は分類精度である．青色の線で示した”Ideal”は認識対象の画像と同一の距離で撮影された画像を訓練に用いた場合の精度，赤色の線で示した”Proposed method”は認識対象の画像と同一の距離の仮想視点画像を訓練に用いた場合の精度，そして緑色の線で示した”Xu’s method”は認識対象の画像と同一の距離の画像を訓練に用いない場合の精度である．”Ideal”ではどの認識対象に対しても同一の距離の画像を訓練に用いているため，最も認識率が高い．”Proposed method”では同一の距離の画像を仮想的に生成して訓練に用いているため，”Ideal”に比べて精度が落ちるが，仮想視点画像を用いない”Xu’s method”に比べて低下を抑制できている．横軸が 0 の場合を除いて提案手法の精度が Xu らの手法による精度以上となっており，生成した仮想視点画像も併せて訓練に用いることで認識率の低下を抑えることが出来ることを確認した．

横軸が 0 の場合において提案手法が Xu らの手法による精度を下回っている理由としては，この場合に限って認識対象と訓練画像は同一の距離であるため仮想視点画像による認識率の向上は見込めないのに加えて，認識対象の物体のヒストグラムと生成した仮想視点画像から得た異なる物体のヒストグラムとが類似してしまい，誤認識を引き起こすケースが生じてしまうためである．横軸が 3 の場合において認識率が向上していない理由としては，ライトフィールドレンダリングにより生成した画像に欠損が生じてしまっていることが挙げられる．撮影時の位置から仮想視点の位置までの距離 d が大きいほど，図 5 で示したように中心視点から離れた視点で記録された情報が必要となる．そのため d が一定値を超えると記録されていない光線の情報が必要となり，生成画像に欠損が生じてしまう．このため物体部分の LFD を十分に取得できていないことが原因であると考えられる．

5. まとめ

本研究ではライトフィールドカメラを用いて撮影されたデータを用いて，ライトフィールドレンダリングによって仮想視点画像を生成し，従来手法における距離の変化による認識率の低下を抑制する手法を提案した．提案手法の改善点としては，対応することが出来る領域が狭いことが挙げられる．撮影位置からどの程度まで離れた位置の仮想視点を生成できるかはライトフィールドカメラに依存するが，この範囲を拡大することは困難である．そこで，この距離に対して不変な空間への射影を求めることで，既に求めた LFD から新たなデータにおける LFD を計算し，計算量の削減と頑健性の向上を実現することを目標とする．

参考文献

[1] Viola, Paul, et al.: *Rapid object detection using a boosted cascade of simple features.*, Computer Vision and Pattern Recognition(2001).
[2] Dalal, Navneet, et al.: *Histograms of oriented gradients for human detection.*,Computer Vision and Pattern Recognition(2005).
[3] Wu, Bo, et al.: *Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors.*, IEEE(2005).
[4] Bartlett, et al. *Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction.*, Computer Vision and Pattern Recognition Workshop(2003).
[5] Mita, Takeshi, et al. *Discriminative feature co-occurrence selection for object detection.*, IEEE Transactions on Pattern Analysis and Machine Intelligence 30.7 (2008): 1257-1269.
[6] Zhu, Qiang, et al. *Fast human detection using a cascade of histograms of oriented gradients.*, Computer Vision and Pattern Recognition(2006).
[7] Wu, Bo, et al. *Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors.*, International Journal of Computer Vision 75.2 (2007): 247-266.
[8] Krizhevsky, Alex, et al.: *Imagenet classification with deep convolutional neural networks.*, Advances in neural information processing systems(2012).
[9] He, Kaiming, et al.: *Deep residual learning for image recognition.*, Proceedings of the IEEE conference on computer vision and pattern recognition(2016).
[10] Xu, Yichao, et al.: *Light field distortion feature for transparent object classification.*, Computer Vision and Image Understanding 139 (2015): 122-135.
[11] Levoy, Marc, et al.: *Light field rendering.*, Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. ACM(1996).
[12] Brox, Thomas, et al.: *Large displacement optical flow: descriptor matching in variational motion estimation.*, IEEE transactions on pattern analysis and machine intelligence 33.3 (2011): 500-513.
[13] Xu, Yichao, et al.: *Transcut: Transparent object segmentation from a light-field image.*, Proceedings of the IEEE International Conference on Computer Vision(2015).