計測自動制御学会東北支部 第 303 回研究集会 (2016.7.15)
資料番号 303-1

視覚的顕著性を用いた オクリュージョンに頑強な物体領域の抽出

Occlusion-Robust Extraction for Object Regions using Visual Saliency

○海沼旭，間所洋和，佐藤和人，下井信浩

○ Asahi Kainuma, Hirokazu Madokoro, Kazuhito Sato and Nobuhiro Shimoi

秋田県立大学　機械知能システム学専攻

Department of Machine Intelligence and Systems Engineering, Akita P_{rec}fectural University

キーワード： MAV (Micro Air Vehicle)，オクリュージョン (Occlusion)，領域分割 (Segmentation)，顕著性マップ (Saliency maps)，GrabCut

連絡先： 〒 015-0055　秋田県由利本荘市土谷字海老ノ口 84-4
秋田県立大学 システム科学技術研究科 機械知能システム学専攻 脳情報工学研究室
海沼旭，E-mail: M18A008@akita-pu.ac.jp

1. 緒言

機械技術の進展や演算装置の高性能化に伴い，多種多様な機能を有するロボットが開発されている．特に工場や極限環境では，繰り返し動作の単純作業だけでなく，人間にとって危険な仕事や不可能な作業をロボットが代行している．災害時には，MAV（Micro Air Vehicle）や UGV（Unmanned Ground Vehicle）などの活躍も期待されている．これらのロボットに更に求められているのは，適用する環境に合わせて人間がロボットを管理し命令を与えるだけでなく，ロボット自身が自律的かつ合目的に環境を理解し行動する能力である．環境の意味的な理解には視覚機能の果たす役割が大きく，計算機を用いて物体やシーンをカテゴリとして分類し認識することが，研究課題の一つになっている [1)]．一方，物体認識においては画像中の対象物体が属するカテゴリの判別基準は曖昧であり，すべての画像に対して同様の判別基準を用いることは現実的に困難である．また，背景が複雑な画像でも，画像中の物体の判別が著しく困難になる．更に，同一カテゴリに属する物体であっても色や形状が異なる場合や，同一物体であっても視点や照明などの撮影条件に依存して見え方が異なる場合，物体の認識は極めて困難である．このような課題に対して，カテゴリ内の物体条件や撮影条件の変動に頑強な特徴記述法，物体検出法，物体やカテゴリの表現法，識別器等が研究されてきた [1)2)]．コンピュータビジョンにおける物体認識では，公開データセットによる性能評価が主流を占めている．これらのデータセットは WEB 上で入手可能で，一般物体認識の研

究の標準的なベンチマークデータとなっている．しかしながら，スナップ写真に近い画像により構成されており，対象物体が中央に存在する傾向にある．また，識別する物体のカテゴリ数が事前に決められている．これに対して，ロボットビジョンにおいては，対象となる物体がシーンの中央に存在するとは限らない．更に，実環境では多種多様な物体が存在し，手前の物体が背後にある物体を隠して見えない状態であるオクリュージョンが発生する．したがって，学習データセットの構築におけるオクリュージョンの取り扱いや，画像の取得と教師信号の付加に伴う負荷の軽減や，画像を取得するロボットの選定が検討課題になっている．また，物体認識における，従来研究では，物体の抽出と認識が個別に研究されていた [1]．認識の前処理である物体抽出の研究では，高精度に抽出するため，教師あり学習による手法が盛んに研究されている [1]．一方，教師あり学習による物体認識では，オフィスや家庭などの一般的な環境を対象とした場合，膨大な量の学習データが必要となる．また，人手による教師データの作成が必要となるため，学習データの構築に多大な負荷が生じる．そのため，学習を必要としない物体の抽出が注目されている．

本論文では，MAV を利用して物体認識を行うことを目的とする．ロボットビジョンとしての MAV は，アクティブビジョン特性に優れ，対象物体をあらゆる角度からセンシングし，画像データとして取得することができる．この多視点性は，物体認識の問題の一つであるオクリュージョンに対して有効であることに，本研究では着目した．更に，取得した画像を対象に，顕著性に基づく物体の分割の観点から，学習を必要としない物体領域抽出法を提案し，その有効性を評価する．評価実験では MAV で構築したオリジナルデータセットを用いた性能評価により，提案手法の性能を多面的に検証する．また，MAV による画像データの取得の有効性も考察する．

2. 関連研究

領域分割による方法では，オクリュージョンがある場合や，形状が複雑で領域分割に誤差が多く，精度が低い場合には，認識精度が著しく低下する．Schmid ら [3] は，領域分割を行わずに，画像中の局所的な特徴の組み合わせによって，画像の照合を行う方法を提案した．具体的には，始めに Harris 特徴点オペレータ [4] によって，画像中から 100 点程度の特徴点を選定する．次に，各点の画素値や微分値などを特徴ベクトルとし，それらの集合によって 1 枚の画像を特徴付けた．照合は，未知の画像に対して，同様に特徴ベクトルの集合を求める．そして，学習画像の特徴ベクトルを探し，類似している学習画像に対して，投票を行う．この時，特徴点数間の相対的位置関係を考慮することによって，無駄な投票を防ぎ，最終的に最も多くの投票を集めた学習画像に一致したとみなす．特徴点自体は 1 画素の座標に過ぎないが，特徴点付近の画素から得ることができる高次元特徴ベクトルで特徴付けることによって，点の集合による画像の表現が可能となった．Lowe も同様の考えによって独自に手案した特徴点抽出とその記述法を合わせた SIFT(Scale Invariant Feature Transform) を用いて，オクリュージョンのあるシーンにおける物体認識を実現している [5]．

MAV を利用した物体認識として，Sinisa ら [6] は MLDA（Multiscale Linear Discriminant Analysis）を用いた手法を提案した．この手法は，空と地面の境界線の検出に優れており，TSBN（Tree-Structured Belief Networks）の確立的表現を用いて境界線を決定する．始めに，MAV で取得した画像の特徴点を検出する．その後，分割した地面の領域に対して，コンテクストの概念に加えて，CWT(Complex Wavelet Transform) と HSI 色空間を用いることで，関心対象を高速に検出する．MAV を利用して取得した 20 枚の単一物体を対象とした画像に対しての評価実験は，物体の検出精度が 87%であった．しかしな

Fig. 1 Our platform MAV (AR.Drone2.0; Parrot Corp.).

がら，対象とした画像は，すべて屋外のもので，多様な人工物から構成される屋内における有効性は未知である．また，２物体を対象とした評価実験は存在したが，オクリュージョンの有無には触れられていない．

3. MAV

本研究では，フランス Parrot 社の AR.Drone2.0 をプラットフォームとして用いた．AR.Drone2.0 の概観を Fig.1 に示す．AR.Drone2.0 は比較的安価に購入でき，SDK（Software Development Kit）の開発が盛んに行われているため，利便性や拡張性に優れている．本体のサイズは，全長 515mm，全幅 520mm，最高速度は 5m/s であり，室内で運用するには，大きさ，速度ともに適している．視覚系には，前方に 720p の HD カメラ，下方に対地速度測定用の QVGA カメラが搭載されており，フレームレートは前者が 30fps，後者が 60fps である．本実験では前方の HD カメラのみを使用する．

本研究では，Puku らが開発した SDK の CV-Drone を用いた．CVDrone は OpenCV を活用した画像処理機能の装備が特徴となっている．なお，OpenCV とはコンピュータで画像や動画を処理するための様々な機能が実装されているオープンソースのコンピュータ・ビジョン・ライブラリである．AR.Drone2.0 の実現可能な飛行パターンは前進，後進，横移動，旋回移動，上昇，下降移動，ホバリングである．これらの動作から多視点画像を得るための飛行パターンについて検討し，直進飛行，水平方向の S 字飛行，垂直方向の S 字飛行，旋回飛行の 4 パターンを抽出した．評価実験では，これらの中から，多視点画像を得るために最適な飛行パターンを選定する．

4. 提案手法

本研究では，ロボットビジョンを用いて物体領域の抽出を行う．物体領域を抽出する手順は以下のように細分化される．

(1) 画像中の物体を含む関心領域の抽出

(2) 関心領域の分割による物体領域の抽出

ロボットが自律的に複数の物体を認識するためには，自動で対象となる物体とその背景，つまり前景領域と背景領域に分割しなければならない．さらに，前景と背景の分割には，予め物体の位置や大きさを関心領域として抽出する必要がある．

本論文では，複数の物体を含む関心領域を検出し，物体領域に分割する手法として，Itti ら [7] によって提案された顕著性マップと Carsten ら [8] によって提案された GrabCut を組み合わせた手法を提案する．提案手法の処理手順を Fig.2 に示す．始めに，顕著性マップを用いて，原画像に対して注視点を検出する．顕著性マップは画像中に存在する視覚的注意をひく領域を自動的に検出する手法である．画像中に存在する物体は，看板や障害物のように，視覚的注意を引くため，顕著な領域は物体に表出しやすい．そのため，注視点を検出することで，物体が存在する位置を特定する．顕著性マップでは，輝度，色，方向性の 3 種類の成分を基にマップを作成し，注視点を検出する．

続いて，検出した注視点を基に，物体を包含した関心領域を抽出する．関心領域の抽出では，

対象物体の大きさに合わせた領域となるように，回転や拡大，縮小に頑強な SIFT を用いた．最後に，抽出した関心領域を GrabCut に付与することで物体領域を抽出する．

本手法では，関心領域の内側を物体らしい領域，外側を背景らしい領域として，GrabCut に疑似的教師データを付与する．そのため，学習なしの領域分割が可能となる．本手法の特徴は，顕著性マップを用いて注視点を検出し，自動的に物体の位置が検出できる点にある．また，顕著性マップで複数回処理することにより，複数の物体が存在する位置を注視点として検出が可能となる．

ここでは特に関心領域抽出法について詳しく述べる．始めに，顕著性マップより検出した注視点 (X_{SPs}, Y_{SPs}) と SIFT の特徴点 (X_{sift}, Y_{sift}) のユークリッド距離 D を次式により算出する．

$$D = \sqrt{(X_{sift} - X_{SPs})^2 + (Y_{sift} - Y_{SPs})^2}. \quad (1)$$

続いて，D が小さい特徴点を順に選択する．ここで，選択する特徴点数の比を m とする． 選択した特徴点群より最大座標 $(Xmax, Ymax)$ と最小座標 $(Xmin, Ymin)$ 内の SIFT の最大スケール値 S を求める．以上より，$(X_{min} - S, Y_{min} - S)$ から $(X_{max} + S, Y_{max} + S)$ までの範囲の格子領域を設定する．

5. 評価実験

MAV の移動や環境の変動に伴い，視野画像は時々刻々と変化する．これに伴い，対象物の見え方も多種多様に変化する．そのため，実環境において，対象物の見え方の変化やオクリュージョンの発生に対して頑強な物体抽出が望まれる．本章では，AR.Drone2.0 を用いて視野画像を取得し，見え方の変化や，オクリュージョンに対する提案手法の性能を評価する．

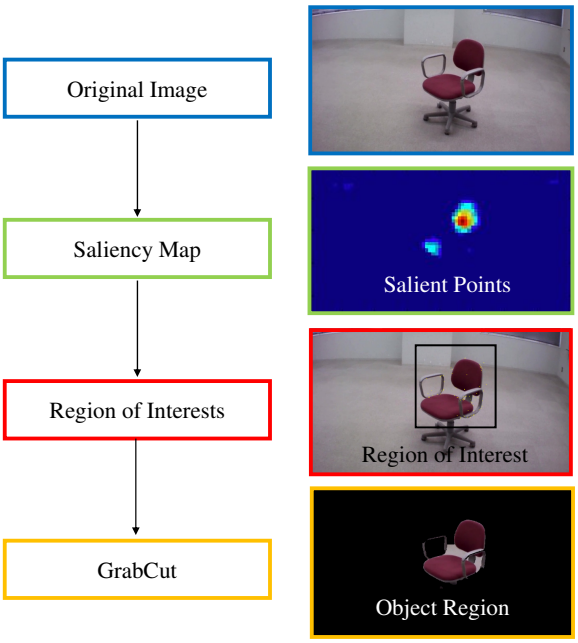

Fig. 2 Processing Procedure.

5.1 実験条件

AR.Drone2.0 の飛行パターンは旋回飛行のみとした．この飛行パターンが，多視点性を最も有効に生かすことができると考えたからである．しかしながら，安定して旋回飛行の操作をすることは難しいため，本研究では手持ちの模擬飛行により，視野画像を取得した．旋回飛行時の円の半径は 1.5m，飛行の高さは 1.5m とした．

実験は，著者らの所属する大学の会議室で実施した．実験環境と対象物体を Fig.3 に示す．本環境は，ゼミや会議に用いられるため，物体は椅子や机がほとんどであり，背景の複雑度が低いことから，本実験に適していると考えられる．また，窓からの太陽光やガラスの反射光を防ぐために，ブラインドを下ろした状態で映像データを取得した．本実験では，椅子と机の 2 つの物体を抽出の対象とした．Fig.3 の objects の部分に，椅子のみを設置した場合，机のみを設置した場合，椅子と机の両方を設置した場合の 3 パターンを設定した．この組み合わせを，順にパターン A，パターン B，パターン C として，Table1 に示す．実験画像としては，取得した視

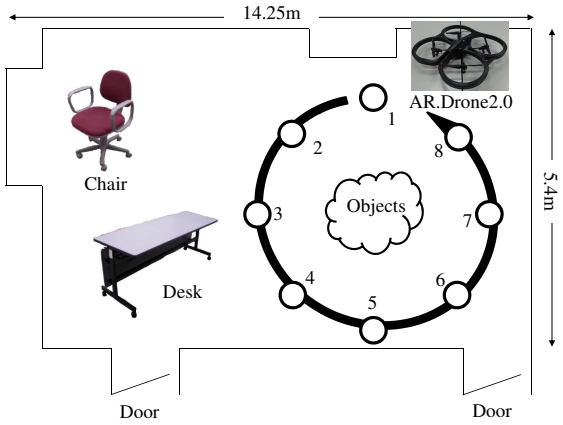

Fig. 3 Experiment environment with the target objects and flight trajectory.

Table 1 Three patterns of obtained datasets.

Pattern	Objects	Occlusion
A	chair	n
B	desk	n
C	chair and desk	y

野画像の中から均一の間隔で，8 枚の画像を切り出した．なお，Fig.3 において，白丸が撮影位置を表し，数字が選定された画像の番号を表す．また，GT は著者がペンタブを用いて，手動で作成した．

5.2 注視点検出結果

顕著性マップにおける注視点検出結果では，パターン A，パターン B は 8 枚の全画像に対して，物体上に注視点を検出した．しかし，パターン C は 8 枚の画像に対して，6 枚は両物体上に注視点を検出し，残りの 2 枚は一部の物体上のみに注視点を検出した．後者の画像を Fig.4 に示す．Fig.4(b) の No.6 の画像は机の下部に顕著性が高く現れており，後方の椅子は顕著度が低くなっている．また，(d) の No.7 の画像は椅子のバックシートとシートが最も顕著度が高くなっているが，机の下部にも高い顕著度を示している．

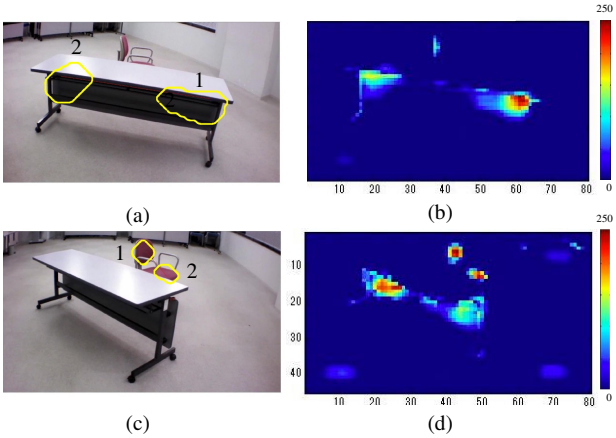

Fig. 4 Detection results of attentional regions: (a) high saliency regions of the sixth image, (b) SM of the sixth image, (c) high saliency regions of the seventh image, and (d) SM of the seventh image.

5.3 領域分割結果

分割精度は，次式に示す適合率，再現率，F 値の P_{rec}，R_{eco}，F_{mea} を用いて評価する．

$$P_{rec} = \frac{N \cap C}{N}, \tag{2}$$

$$R_{eco} = \frac{N \cap C}{C}, \tag{3}$$

$$F_{mea} = \frac{2(N \cap C)}{N + C}, \tag{4}$$

ここで，N は物体として抽出した領域の画素数，C は GT 領域の画素数である．また，先行研究 [9] より SIFT の選択点数比は m=1/6，GrabCut の更新回数は n=9 とした．

各パターンの抽出率の F_{mea} を Fig.5 に示す．椅子のみを対象とした実験では，8 枚の画像に対して，すべての画像で物体上に注視点を検出した．抽出率の平均は，F_{mea} が 55.84%であり，P_{rec} と R_{eco} は 88.35%と 46.99%であった．また，机のみを対象とした実験では，8 枚の画像に対して，すべての画像で物体上に注視点を検出した．抽出率の平均は，F_{mea} が 42.25%であり，P_{rec} と R_{eco} は 67.17%と 39.18%であった．そして，椅子と机の 2 物体を対象とした実験で

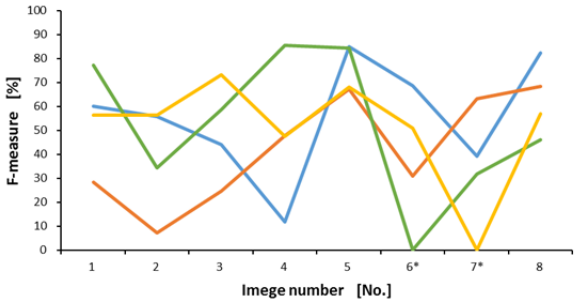

Fig. 5 Segmentation accuracies in respective patterns (blue line: Pattern A, red line: Pattern B, green line: Pattern C of chair, and yellow line: Pattern C of desk. The accuracies dropped to zero signify false extraction of attentional points.

は，8 枚の画像に対して 6 枚の画像は，両物体上に注視点を検出した．しかし，残りの 2 枚の画像は 1 物体上のみに注視点を検出した．両物体上に注視点を検出した画像に対する抽出率の平均は，F_{mea} が 62.48%であり，P_{rec} と R_{eco} は 60.61%と 70.14%であった．椅子のみと机のみの F_{mea} は 57.99%と 58.37%であった．P_{rec} と R_{eco} は，椅子のみが 53.97%と 74.64%，机のみが 67.77%と 59.87%であった

6. 考察

始めに，注視点の検出結果に関して考察する．単一物体を対象とするパターン A，パターン B は，両者とも物体上に注視点が検出できている．しかしながら，複数物体を対象とするパターン C においては，いずれか一方の注視点の検出に失敗している画像が存在している．両画像ともオクリュージョンが発生しており，机の後方に椅子が存在する関係になっている．一方，残りの 6 枚の画像は両物体上に注視点を検出しており，MAV による上空からの多視点性を有する画像取得の有効性を示唆する結果になっている．

続いて，パターン C におけるオクリュージョンの有無による抽出精度の変化について考察する．椅子の抽出結果は，Fig.6(a)(b) に示す No.4 の結果が最も高い精度となった．この画像はオクリュージョンが発生していないため，椅子単体で抽出されており，高精度に結びついたと考える．そして，Fig.6(c)(d) のオクリュージョンが発生している No.7 の画像が最も低い精度となった．これは，机が椅子と共に物体領域として抽出されているため，オクリュージョンの発生による抽出精度の低下が考えられる．

しかしながら，パターン C における机に関しては，オクリュージョンが発生している No.3 の

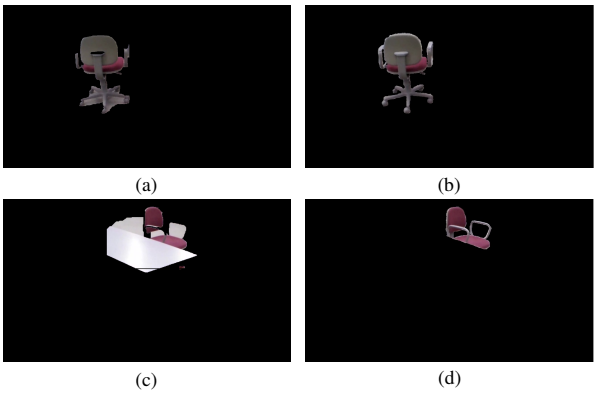

Fig. 6 Segmentation results and GT for the chair: (a) result of the fourth image, (b) GT of the fourth image (c) result of the seventh image, and (d) GT of the seventh image.

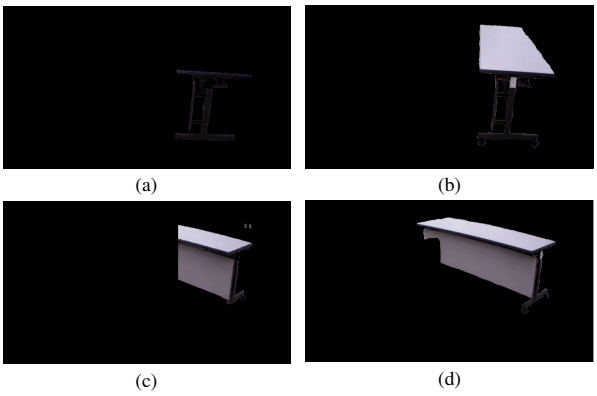

Fig. 7 Segmentation results and GT for the table: (a) result of the third image, (b) GT of the third image (c) result of the fourth image, and (d) GT of the fourth image.

画像の抽出精度が最も高い．一方，オクリュージョンが発生していない No.4 が最も抽出精度が低くなっている．これらの結果を Fig.7 とに示す．最も精度が高い Fig7(c)(d) は机の半分のみ抽出している．これは，GrabCut で関心領域を抽出する際に，机の半分のみを関心領域として抽出しており，左側の椅子のオクリュージョンは包含できていない．また，最も精度が低い Fig.7(a)(b) は，机の淵である黒い部分しか抽出されなかった．これは，机の表面の色と，床の色が類似していることや関心領域の設定方法に起因していると考えられる．よって，パターン C に関しては，相対的に画像内において占有比率の小さい椅子は，オクリュージョンに対するロバスト性の高い結果が得られているものの，画像全体を包含する机は，部分的に抽出されることから，GT との比較による精度に関しては，椅子とは相反する傾向になっている．

一方，本手法では，GrabCut を適用する範囲は相対的に決めているため，MAV が対象物体から遠のいて，視野内における占有割合が小さくなることで，関心領域に関する問題を回避することができるが，椅子に関しては，相対的に小さくなるため，抽出精度に影響を与えてしまう．よって，関心領域の比率が，適応的に設定できるメカニズムの導入が求められる．

7. 結論

本研究では，オクリュージョンに対して頑強な物体認識の実現を目的として，MAV の多視点性を利用した物体抽出法を提案した．本手法は，視覚的顕著性に基づく物体の検出と分割の観点から，顕著性マップと GrabCut を組み合わせた学習を必要としないメカニズムを特徴とする．提案手法において，オクリュージョンが発生している状態で，抽出精度が低下した．しかし，MAV のアクティブビジョンの優位性を，有効に利用することでオクリュージョンを回避することができ，抽出精度が向上した．よって，MAV を利用した物体認識は，オクリュージョンに対するロバスト性が高いことを示した．

今後の課題としては，MAV が自律飛行で視野画像を取得できるようにプログラムを作成することや，MAV を改造して視野を広げること，物体領域の抽出精度を向上させるためにプログラムのパラメータを調整することなどが挙げられる．

参考文献

1) K. Yanai, “The Current State and Future DiR_{eco}tions on Generic Object R_{eco}ognition,” *IPSJ Computer Vision and Image Media*, vol.48, pp.1–24, 2007.
2) T. Okumura, T. Takiguchi, and Y. Ariki, “Generic Object R_{eco}ognition Using CRF by Incorporating BoF as Global Features,” *Meeting on Image R_{eco}ognition and Understanding*, pp.95–102, 2009.
3) Cordelia Schmid, Roger Mohr, “Local Gray-value Invariants for Image Retrieval” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 19, No.5, pp.530-535 1997.
4) C. Harris and M. Stephens, “A Combined Corner and Edge Detector,” *Proc. Alvey Conference*, pp.147–152, 1988.
5) D.C. Lowe, “Object R_{eco}ognition from local scale-invariant features,” *Proc. IEEE International Conference on Computer Vision*, pp.1150–1157 1999.
6) S. Todorovic and M.C. Nechyba, “A Vision System for Intelligent Mission Profiles of Micro Air Vehicles,” *IEEE Trans. Vehicular Technology*, vol.53, no.6, pp.1713–1725, 2004.
7) L.Itti, C.Koch and E.Niebur,“A Model of Saliency-based Visual Attention for Rapid Scene Analysis”,IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.20, No.11, pp.1254-1259, 1998.
8) C. Rother, V. Kolmogorov, A. Blake,““GrabCut”,Interactive Foreground Extraction using Iterated Graph Cuts”,IEEE Trans. ADOBE SYSTEMS INCORP. 2002.
9) A. Yamanashi, H. Madokoro, Y. Ishioka, and K. Sato, “Visual Saliency Based Segmentation of Multiple Objects Using Variable Regions of Interest,” *Proc. 14th International Conference on Control, Automation and Systems*, pp.88–93, 2014.