

# 単眼カメラによる手指の平面タッチ判定とその応用

月岡 陽人<sup>†</sup> 中島 克人<sup>†</sup>

東京電機大学<sup>†</sup>

## 1 はじめに

近年、画面を直接タッチする操作方法はスマートフォンやタブレット、あるいは、大型タッチパネルの普及により一般的となってきたが、タッチセンサ無しで物体表面上の手指のタッチ判定が実現できると様々なインターフェースとその応用が可能となる。プロジェクタとカメラ、あるいはカメラと深度センサを用いて壁や机をタッチパネル化する仮想タッチスクリーン[1]の提案が行われてきたが、複数の装置が必要で、それらの配置の制約も強かった。今回我々は単眼カメラのみを用いて、深層学習に基づく手指のタッチ検出を実現した。また、その応用として、手軽な楽曲制作環境の提供を目的に紙に印刷した鍵盤を用いる仮想キーボードを試作し、評価したのでその結果を報告する。

## 2 提案手法

本手法は 640×480 pixel 程度の解像度の単眼カメラで手指タッチ面の 30 cm程度以内から撮影した映像を入力とし、深層学習による物体検出手法を用いてタッチの検出を行う。

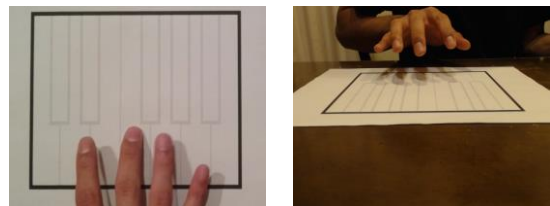
今回は、白紙に印刷された 1 オクターブの鍵盤(紙鍵盤)を平らな机上においてタッチ面とし、鍵盤キーに対応した音階を出力する仮想キーボードを試作した(図 1 参照)。以降に処理の流れ、タッチ検出の詳細などについて記す。

### 2.1 処理の流れ

- (1) カメラ映像により鍵盤領域の輪郭抽出を行い、その輪郭に基づき鍵盤各キー位置を記憶する(カメラキャリブレーション)。
- (2) カメラ入力各フレームに対して、深層学習による物体検出手法によりタッチ検出を行う。
- (3) タッチ検出がある場合、全ての検出矩形の下辺の中心座標に対応する鍵盤キーをそれぞれ求め、そのキーの音階を出力する。
- (4) (2)~(3)を繰り返す。

### 2.2 鍵盤領域の検出

手指がタッチした座標に対応する鍵盤キーを決定するために鍵盤領域のキャリブレーションを行う。



(a) 紙鍵盤 (b) カメラ画像  
図 1. 仮想キーボードとそのタッチ検出

まず、紙鍵盤の元画像をテンプレートとし、テンプレート内で鍵盤キーに対する音階の割り振りを行う。次に図 1(b)のようなカメラ入力画像を 2 値化し、白色の一番大きな面積の領域を紙全体と見なし、更にその内部の黒色の枠を紙鍵盤の輪郭と見なす。そして、その輪郭の四隅を紙鍵盤テンプレートの四隅に対応させるように射影変換行列を求める。以降、カメラ入力上のタッチ座標を射影変換してテンプレート内の鍵盤キーを求める。

### 2.3 タッチ検出

画像の入力から検出までを end-to-end で学習でき、高い検出精度が期待できる深層学習手法である YOLOv3[2]と SSD300[3]の 2 種類を用いてタッチ検出を実装した。

学習には、左右それぞれ 5 本の指で、キーボードで演奏するような手の形で白色紙上をタッチした際の画像を用いる。タッチの際の白色紙上にできるわずかな影を伴った指の形が学習されることを狙う。

## 3 実験と評価

提案手法の有効性を検証するために、図 1(b)のようにその手指の前方から撮影した画像および映像によりタッチ検出の精度を計測した。

### 3.1 カメラと紙鍵盤の配置および照明

単眼カメラ logicool Pro webcam C920t を鍵盤の外枠が全て画角内に収まる紙鍵盤の上辺から約 18 cm の距離に置き、俯角 0° および 20° で撮影を行う。照明には白熱球を使用する。紙には白鍵と黒鍵をドからシまでの 12 鍵印刷した。それぞれのキーの横幅は実際のピアノに合わせ白鍵を 23 mm、黒鍵を 11 mm とした。学習モデルは白紙上のタッチ画像を用いて学習するため、紙鍵盤のキー境界は検出の障害にならないが目で認識はできる薄い灰色(輝度値 195)とした(図 1(a)参照)。

### 3.2 学習用データセット

3.1 に示す撮影条件の下、14名のタッチする手指を直上および左右の3パターンからの光源照射で撮影を行い、タッチしている指の第1関節付近から指先の影を最小限含む範囲までを矩形でアノテーションしてデータセットとした。内訳を表1に示す。学習の際には、YOLOv3とSSD300の入力サイズである320×320, 300×300 pixelに合わせて、データセット画像をアスペクト比は変えず余白はパディングしてリサイズする。

表1. データセット内訳 (枚)

	学習	検証	テスト	計
データ数	1,720	191	478	2,389

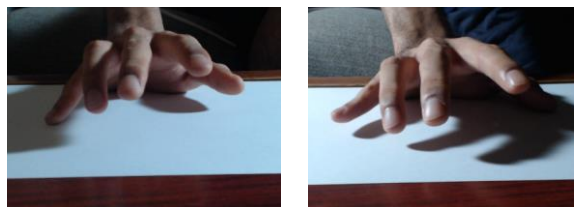
### 3.3 学習および実験結果

まず、それぞれの検出器が学習を行えているかを確認するため、表1に示す学習用データと同条件で撮影されたテスト画像478枚を用いて適合率と再現率の平均を算出した(表2参照)。確信度はいずれも0.6以上とし、適合率は物体検出用の大規模データセットを提供するCommon Object in Contextの形式に倣って、 $IOU \geq 50$ , 75それぞれについても求めた。

表2. 各検出器での精度比較

検出器名	AP	AP <sup>50</sup>	AP <sup>75</sup>	Recall
YOLOv3	0.74	0.77	0.80	0.53
同上(親指)	0.74	0.74	0.78	0.51
同上(小指)	0.39	0.39	-	0.13
SSD300	0.89	0.90	0.92	0.24
同上(親指)	0.96	0.96	0.97	0.22
同上(小指)	1.0	1.0	1.0	0.15

YOLOv3, SSD300共に適合率が高い数値を示しているが、どちらも再現率が低く、タッチの検出ミスが多いことが分かる。これは図2のように、小指が他の指に比べてカメラから見える形状が異なることや、短い親指が人差し指に一部隠れる事と照明位置によっては影により暗くなり過ぎる事等が原因と思われる。表2に親指と小指のタッチ画像それぞれ11枚と6枚に対する精度も示す。



(a) 小指タッチ画像 (b) 親指タッチ画像  
図2. 検出ミスの例

次に、実際に鍵盤上で曲を演奏する際の運指でもタッチ検出が行えるかを確認するため、演奏映像からのタッチ検出を行い、適合率と再現率を求めた(表3参照)。曲は1オクターブ内でメロディーが演奏で

表3. 実験映像から各検出器での適合率と再現率の平均

検出器名	AP	Recall
YOLOv3	0.97	0.90
SSD300	0.88	0.71

きる「きらきら星」とし、テンポは四分音符で毎分60回(J=60)とした。演奏者は3名である。

テスト画像と異なり、仮想キーボードの室内での利用環境として想定される天井からの複数の蛍光灯照明下で、スマートフォンを前方机上に置いたようなほぼ水平方向からの撮影を行った。テスト画像による検証時に比べ適合率、再現率共に向上している。しかし、小指と親指に限定し、適合率と再現率を求めると、両検出器共に適合率および再現率が0.5以下となっており、曲の演奏中の運指においても小指と親指のタッチ検出精度に課題を残した。

次に、同じ演奏映像から、演奏者がタッチした紙鍵盤上のキーと仮想キーボードが発出した音階を比較し、正しい音階を発出できた割合を算出した。その結果、YOLOv3では0.80, SSD300では0.71となった。正しく音階を発出できなかった殆どが、タッチ検出矩形の下辺の中心座標と実際に指がタッチしている位置とが大きくずれていたためであった。

なお、CPUがi5/2.90GHz, 搭載GPUがGeForce RTX 2060, メモリが15GBの計算機でのタッチ検出処理時間は、映像フレームが入力されてから当該キーの音階の判定までYOLOv3では約50msec, SSD300では約71msecであった。

## 4 まとめ

本稿では単眼カメラによる映像を入力としたタッチ検出の実現と、応用として紙で作られた鍵盤を用いた仮想キーボードの試作を行った。実験の結果、一般物体検出器を用いてタッチ検出を実現できること、および、仮想キーボードのような応用システムが容易に実現できることを示した。

今回の実験での未検出の多くは小指や親指であり、タッチ時のカメラから見える指の形状の違いや他の指に部分隠れすることがその原因であると分かった。従って、タッチ検出の精度向上には、学習用データセットにこれらのパターンの画像を増強する必要がある。また、タッチ位置を検出矩形の下辺中心と決めず、検出矩形内の指の形状を考慮することでタッチ位置の精度向上も図りたい。

## 参考文献

- [1] Takashi Homma, Katsuto Nakajima, "Virtual Touch Screen 'VIRTOS' -Implementing Virtual Touch Buttons and Virtual Sliders using a Projector and Camera," VISAPP2014, vol4, pp.34-43, 2014.
- [2] Joseph Redmon and Ali Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767[cs.CV], 2018.
- [3] Wei Liu, et al., "E-SSD: Single Shot MultiBox Detector," arXiv:1512.02325[cs.CV], 2016.