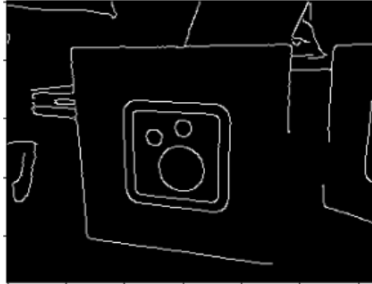


## Extraction de la région d'intérêt

Afin d'extraire la région d'intérêt dans chaque image, nous pouvons utiliser la librairie OpenCV pour détecter la région automatiquement. Pour ce faire, nous procédons de la façon suivante :

### Préparation de l'image

1. Application d'un filtre gris sur l'image
2. Application d'un filtre bilatéral **cv2.bilateralFilter** pour réduire le bruit sur l'image
3. Utilisation de la fonction de détection de bord **cv2.Canny**

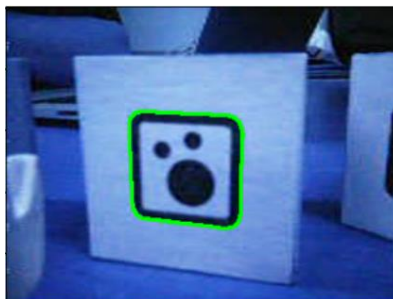


### Extraction de la région d'intérêt

Essentiellement, on doit chercher les régions dans l'image qui ont un contour carré puisque chaque symbole est à l'intérieur d'un cadre noir. On doit effectuer les étapes suivantes :

1. Détecter des contours à l'aide de la fonction **cv2.findContours** (opencv)
2. Filtrer la liste des contours afin d'éliminer les contours qui ne sont pas carrés ou qui n'ont pas une certaine aire. On veut éliminer le plus possible de faux positifs, soit les contours qui ne contiennent pas de symbole.
3. Créer une nouvelle image à partir des coordonnées du contour détecté.

Il est bon de noter que la détection de la région d'intérêt n'étant pas parfaite, il se peut qu'aucune région carrée ne soit détectée dans une image. À l'opposée, il se peut que des régions carrées erronées soient détectées.



## Préparation de la région extraite

Maintenant que la région d'intérêt est extraite, on doit faire quelques transformations avant de l'ajouter à l'ensemble de données (dataset).

1. Normalisation des dimensions de l'images en 40 x 40. Ces dimensions était un bon compromis pour diminuer la quantité de caractéristiques sans perdre trop de détail.
2. Application d'un filtre gris sur l'image.
3. Application d'un flou sur l'image pour faire disparaître les détails.



À la fin, il est recommandé d'afficher dans une grille un échantillon aléatoire d'images. Cela permet de vérifier l'extraction des régions d'intérêt d'images prises au hasard et vérifier qu'elles sont bien étiquetées. Il se pourrait que certaines régions détectées soient erronées mais leur taux est assez faible pour avoir peu d'impact sur l'ensemble d'apprentissage.

