computer graphics & vision

Inffeldgasse 16/2, A-8010 Graz

www.icg.tu-graz.ac.at

TUG

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa

Todos los derechos reservados
Eskubide guztiak erreserbatu dira

NON-MAXIMA SUPRESSION

Mikel Diez Buil

Inst. for Computer Graphics and Vision
Graz University of Technology, Austria

Technical Report
ICG–TR–xxx
Graz, August 29, 2011

contact: Mikel Diez zeixote@gmail.com

Abstract

Non-Maxima Suppression is a very important part on the object detection process. When searching for objects in and image several points are usually found as objects but some of them are not really objects, Non-Maxima Suppression (NMS) consists in select which of those maximas are really objects and suppress those that are not. In this thesis different Non-Maxima Suppression for Hough based images methods have been tested, the methods are Gall, Wenzel, Thresh and Islands. Those methods work with Hough images, which are greyscale voting images where white is high probability to be a maxima and black the opposite. The methods have different characteristics and different ways to act depending on the dataset and the used threshold. To compare those methods TUD Pedestrians and TUD Campus datasets were used to obtain the Precision and Recall charts. The different methods had a very different response to the chosen threshold or to the different datasets which in these cases consist in people walking in several places. Gall method is the most robust one, Wenzel and Thresh have almost the same response to the datasets and Islands with a different way to work has a different response than the other ones. The conclusion was that the correct selection of the threshold and the overlap for the different methods and datasets is very important to achieve good results at object detection, what makes creating algorithms for different environments and illuminations much more difficult than for known environments and controlled situations.

Keywords: *Non-Maxima Suppression, Gall, Wenzel, Island, Thresh, Precision&Recall, Hough.*

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

1 Introduction

Object detection is a field in computer vision that is growing everyday. It has many applications in different fields of our life, such us medicine, photography, automation of process, security and of course entertainment. That means that each year we start using more and more of those applications and there are people behind them developing all the algorithms that makes that possible.

Non-Maxima Suppression is a very important part on the object detection process. When searching for objects in and image several points are usually found as objects but some of them are not really objects. In this case those methods are applied to Hough Voting based images. The Hough Images are grey-scale voting images where white means high probability of being object and black the opposite. Non-Maxima Suppression (NMS) consist in eliminate at those images all those points that are detected as objects due to being white (high probability) but they are not really objects or multiple detections in the same object, that are called false positives.

Once the false maximas are suppressed the objects are marked by a bounding box. The bounding box is a green square that is over the object see Figure 1 and that has to fit to it as much as possible so the bounding box size is different for each object depending on its size, what depends a lot in how near to the camera the person is.

There are several ways for Non-Maxima Suppression all of them with their own characteristics and with a different response to the different datasets in this case Gall, Wenzel, Thresh and Island methods were used. The goals of them are the same, avoid multiple bounding boxes for the same object (false positives), avoid merging two objects as one and avoid missing objects at the image without finding it.

Figure 1: LEFT: Image where the algorithm detects two objects where there is only one (using Gall). RIGHT: Image with one missing object(Wenzel).

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

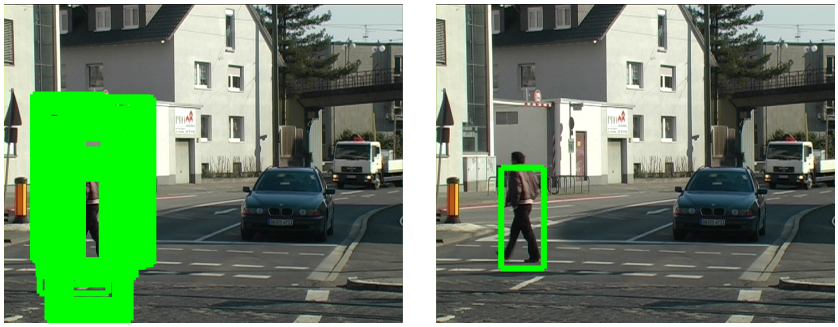

Figure 2: Differences between using or not the BBoxes overlap method.

2 Hough Voting

Hough voting is an essential step in this theses. The objective of it is generate images with detection hypotheses, in those grey-scale images white means high probability of being a object and black the opposite.

This images are obtained by comparing the original image with a code-book and the matches votes for the possible position of the object. The code-book consist in different parts of the object class searched, in this case people. So the code-book will be parts of the human body, hands, feet, heads, and more.

When a match votes it does not vote for the place the match is but for where it thinks the center of the object is. The vote is not a exact point of the image, the match votes with a higher value for the most common places the center to be, but also votes (with less confidence) for other probable places. When it votes it creates an array with the votes and starts accumulating all the votes of the different matches to create the voting image.

At object detection there are usually object at different distances of the camera so there are size differences between the same objects. One option to solve that would be creating at the code-book the same words with different sizes, but when using a code-book one of the most important things is the size of it. So to solve this the algorithm process the same image more than one time changing the scale of it. This results in having several hough voting images for each one of the original images (I will call them here sub-images), one for each of the scales used. This also helps to create the bounding box. If the scale of the object is known, the bounding box can be optimized to de object because the size of it is known.

One of the problems this images have is that the center of the object is blurred what makes difficult for an algorithm to localize the exact center of the object. That would not be a big problem if the objective was finding

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

only one object, because in most cases will be pixel with a highest value at the images, but the objective is to create an algorithm that localizes several people without knowing how many of them are going to appear.

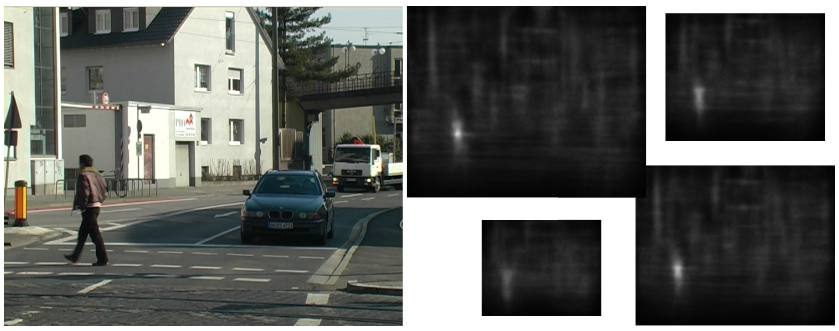

Figure 3: LEFT: The original image of the dataset. RIGHT: The four Hough images (one for each scale, in this case 4) of the original image.

3 Non-Maxima Suppression

As seen in the Figure 1 the center of the object is not an exact point and is blurred around it. Also there are different parts in the images that are not black even if there is nothing there. The goal of Non-Maxima Suppression will be find the true object locations and suppress the others. This makes Non-Maxima Suppression a very important stage in the object recognition because without it it those Hough maps would be useless.

For this experiment two Non-Maxima Suppression methods were used, the Gall method and the Bounding Boxes Overlap method. For the second method three other different methods to find the maximas of the image were used. The maximas are all those points of the image that are over the selected threshold, and once the maximas are found the algorithm selects which of them are real maximas or were just false positives.

The maximas are the different hypothesis of location of the objects, and some of those hypothesis will be right and others will be wrong.

3.1 Gall

This method is a iterative process where when a maxima is found, the nearly pixels are suppressed. The algorithm is a loop that is repeated until no more pixels at the hough images are over the threshold (maximas). The loop consist in the following different steps, find the maxima, interpolate to know

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

how big the bounding box has to be, create the bounding box, remove the maxima and the nearly pixels.

- Find the maxima: At this part of the algorithm the absolute maxima (the highest value) of each of the sub-images are found, storing the value and the position (axis X and Y) in different vectors with length equal to N, where N is the number of scales in the hough images.

- Interpolate: Depending in which sub-image has the highest value the bounding box will have a different size to fit the object as exactly as possible. Even if there are several scales at the hough images usually the objects will not be exactly of the size of one of them and will be between two so an interpolation is necessary. To do this the algorithm interpolates the two highest values obtained in the previous step, those values represent which images are nearer to the scale of the object and ponderating the two values with the known scales a very accurate size for the object can be inferred.

- Build the bounding box: Once the scale of the object is known the bounding box can be built. Being known the center of the object (maxima) what the algorithm does is finding two of the corners of the bounding box, the top-left corner and the bottom-right corner. The program has stored the distance between the center of the object and the different corners for a object of a scale 100%, multiplying this value by the scale obtained in the previous step we obtain the position of the different corners.

- Suppress the maxima: Once the bounding box is created the algorithm proceed to suppress the maxima that created it and its surroundings. This is done by introducing a rectangle of zeros in the sub-images. Depending in which of the sub-images were the maxima the rectangle will be bigger or smaller, bigger if is a big scale and small if its a small one.

3.2 Bounding Boxes Overlap

This method takes the bounding boxes that where found by a previous maxima finding algorithm, in this case Wenzel, Thresh or Island, and selects which of them are the real ones and suppress the ones that are false positives.

This algorithm has two inputs, the bounding boxes of the images and the selected "Overlap". The overlap is a parameter that regulates how much a

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

bounding box can be over an other one, for example an Overlap of a 25% means that if the 25% (or more) of the area of a bounding box is over an other one with more confidence, the one with less confidence will be considered as a non-maxima and will be suppressed. The confidence depends on the value of the maxima that generated the bounding box, a high value is a high confidence and vice versa.

The process that the algorithm follows is the following one:

- Load variables: The first step of the method is loading the position of all the bounding boxes with their confidence and the selected overlap. Then the array with the bounding boxes is reorganized having the bounding boxes organized from the highest confidence to the lowest.

- Process: This algorithm is a iterative process. It takes the first bounding box and start comparing it with the others, if one of the bounding boxes is over other in more than the percentage of the overlap is deleted and the process continues. Then it starts with the next bounding box and repeats the same process until no more bounding boxes are left to compare.

The algorithms used to find the bounding boxes for the Overlap method were the following ones.

3.2.1 Thresh

The Thresh method is the most basic of the methods for finding the maxima used for this comparison. It is a threshold on the hough map that returns all the maximas that are over the selected threshold. The steps are:

- Find Maximas: It keeps all the positions of the pixels that are over the the selected threshold at the hough images in an iterative process that last for the same loops as scales there are.

- Interpolate: It is an iterative process, that is repeated for each of the maximas. In this step the algorithm takes for each of the maximas the value of all the images in that position. Then with the two highest values it makes the same process as in the interpolation of the Gall method.

- Build the bounding box: This part is inside the interpolation loop so at the end of it the result is the bounding box. The way to build it is the same as in the Gall method.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

Due to the simplicity of way it finds the maximas it selects a large amount of hypotheses, which makes this algorithm to have a really high recall, but a very low precision. It also has a high computational cost because the number of bounding boxes that are built.

3.2.2 Wenzel

The Wenzel method is similar to the Thresh method, but instead of finding the global maximas it finds the local maximas of the image. That means that if lots of maximas are together it will only keep the highest one. The steps are:

- Find Maximas: It keeps the position of all the local maximas of the hough images.
- Interpolate/Build the bounding box: The way this is done is the same as the Thresh mode.

Finding the local maximas fix one of the problems of the Thresh method that is selecting two maximas in two contiguous pixels. However it continues having a high recall and a low precision but the computational cost is lower.

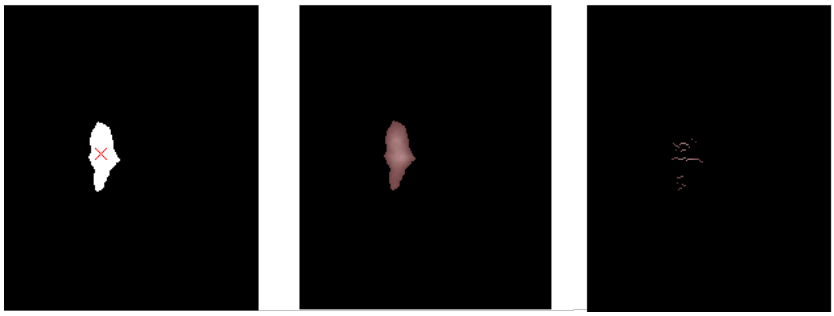

Figure 4: LEFT: Selection using Islands, the center of the cross is the taken point for all the island. CENTER: Selection using Thresh, all is red because all are maximas.RIGHT: Selection using Wenzel.

3.2.3 Islands

This method is a bit different to the previous ones due to the way of finding the maximas. This method uses a more complex way to find the maximas, which consist in creating groups of pixels that are over the threshold and selects only one pixel to represent all of them. The steps are:

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

- Find Maximas: First it eliminates all the pixels of the image that are not over the threshold. Then with the resulting image it creates with all the pixels that are over the threshold regions of pixels that are 8-connected, this regions are called islands. For each island the algorithm finds the maxima and the position of that pixel is what represents the hole island as a maxima and is what is kept.

- Interpolate/Build the bounding box: The way this is done is the same as the Thresh mode.

This method is different to the other two also in its results. The other methods find less maximas if you increase the threshold, but in this case is the opposite. When a low threshold is selected the result is that the islands of different objects got merged because the pixels that separate them now are over the threshold and considered part of the island, and when the threshold is increased the islands break up in different islands. This makes that in a low threshold several objects can be considered one and with a high threshold one object can be considered two. So a low threshold means low Recall and probably low precision, and a high threshold a big recall but also a poor precision. however the objective of this three methods is to have a high recall and the overlap method will be in charge of the precision.

4 Experiment

To evaluate this different Non-Maxima Suppression methods two different datasets of people walking were used. Those datasets are TUD Pedestrians and TUD Campus, which apart of having people walking don't have much more in common. The evaluation of the methods was done with Precision & Recall graphs comparing the different algorithms when using the same threshold.

The do this graphs False Positives and True Positives were used. The False positives are all those bounding boxes that do not belong to a real object, that includes when a bounding box is to big, to small or in the wrong part of a person, and also when is not over a person. The true positives is when the bounding box is over the object with the correct size.

The Recall is how many of the known objects have been found. This is the number of true positives of the image against the total objects of the image. So a high Recall will mean that you have find lots of the objects but without considering how many of the detected objects are not really objects. In the other hand the Precision is how many of the positives are true against the total of detected objects. The precision be low if the algorithm finds lots

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

of the objects but also lots of False Positives. The objective is to achieve a high recall (find all the objects), but also a high precision(do not have false positives).

4.1 TUD Pedestrians

The TUD Pedestrians dataset consist in people walking through different streets at different distances and in different directions. The dataset is formed of 250 images and all of them have at least one person. In most of the cases there are not people partially occluded so their shape is completely visible.

Figure 5: Some TUD Pedestrians Images.

4.1.1 Gall

One of the most important steps at the Gall method is choosing the correct threshold for the algorithm for all the images. When choosing the best threshold for the dataset is important to select the one that achieves both best recall and best precision, but this does not happen in most of the cases. Both are important because it is needed to find all the objects and also is important not to select false positives. In most of the cases one or more than one thresholds have the best recall and is other the one that has the best precision. So to choose the right threshold the best approximation to the best recall and precision has to be chosen.

In the case of Gall method for TUD Pedestrians dataset the threshold selected was 0.30, this means that all the pixels under that value are not considered. To find this several thresholds were used to use Gall at the dataset and this was the one with best relation precision and recall. Also other thresholds were tried, such us 0.35 which has a really low recall or others such as 0,0.1,0.2 or 0.25 that have a low precision. It was interesting to find that even with a threshold of 0 the Gall method does not achieve a 100% recall.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

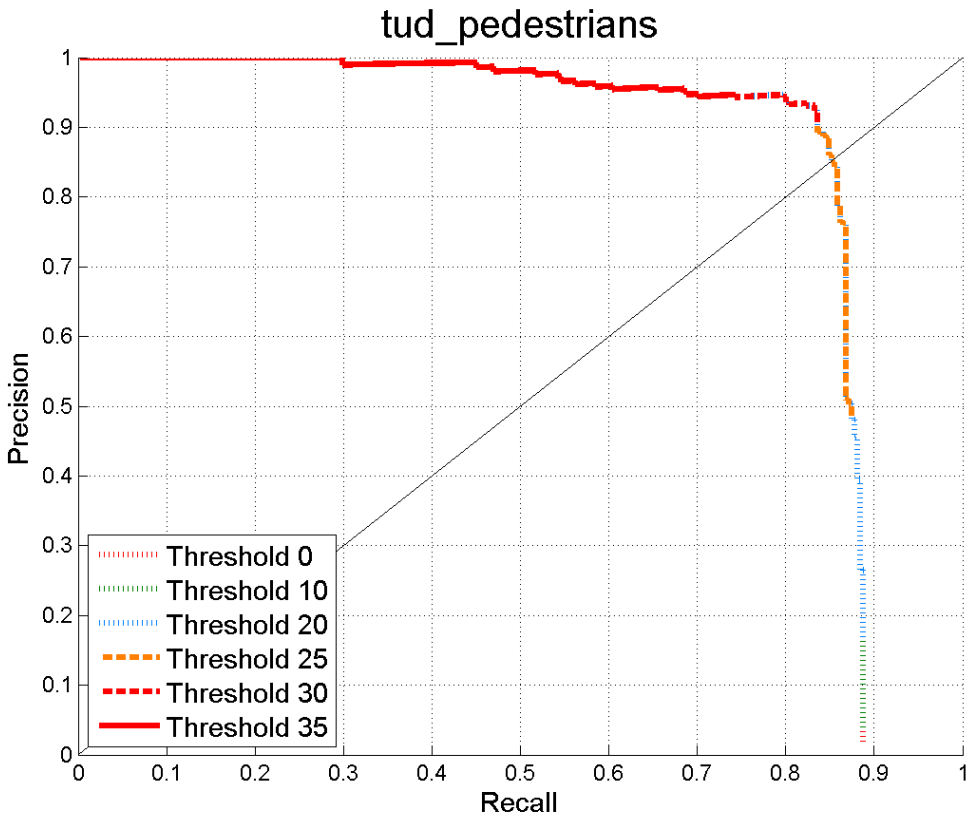

Figure 6: Gall Method for TUD Pedestrians with different thresholds

4.1.2 Gall vs Gall without interpolation

As said before with the Gall method (or the other method to) to find the bounding box the two highest values of each of the iterations are interpolated. As can be seen at the graph of the Figure 7 the response of both ways is very similar, but clearly the Gall method using the interpolation is superior.

The superiority of the interpolated way comes from the fact that when the precision and recall graphs are generated, the different bounding boxes are compared with the ones the dataset provides and if the bounding box is a bit to bigger or smaller its not recognized as correct one even if it is in the right place. The algorithm when using the interpolation reaches a higher Precision, but also a higher Recall and even if there is only a small difference at computer vision having the best algorithm is highly recommended.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

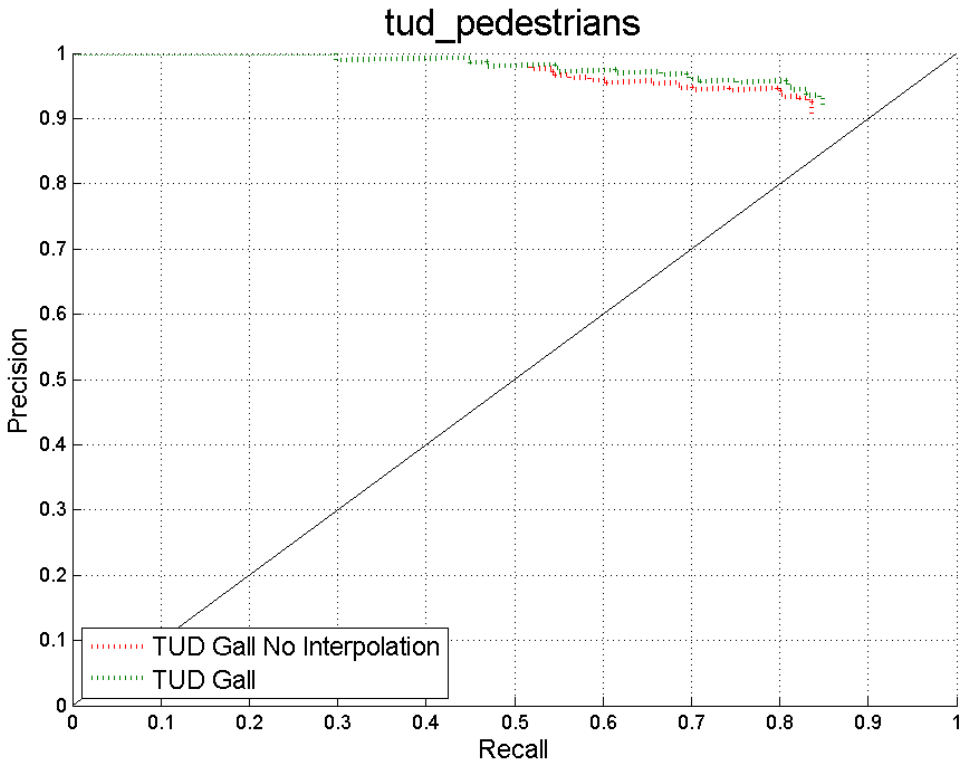

Figure 7: Gall without interpolation vs Gall with interpolation. 30% Threshold

4.1.3 Thresh

Finding the appropriate threshold for this method was a bit more difficult than for the Gall method. In this case the process is splited in two parts, first the bounding boxes are found and then selected which ones are the correct ones. So first was needed to run the algorithm for different thresholds and then those thresholds for different overlaps.

The Thresh method without using the overlap method to select the correct bounding boxes has a very low precision even for high thresholds. With this method and a threshold of 0 the algorithm would achieve a 100% (or very near) because it would create a bounding box for each single pixel of the image. Of all the chosen methods this was by far the one with a higher computational cost because the number of points it selects as maxima, so the minimum threshold used to compared was 40 because of the problems the computer had to use a smaller one.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

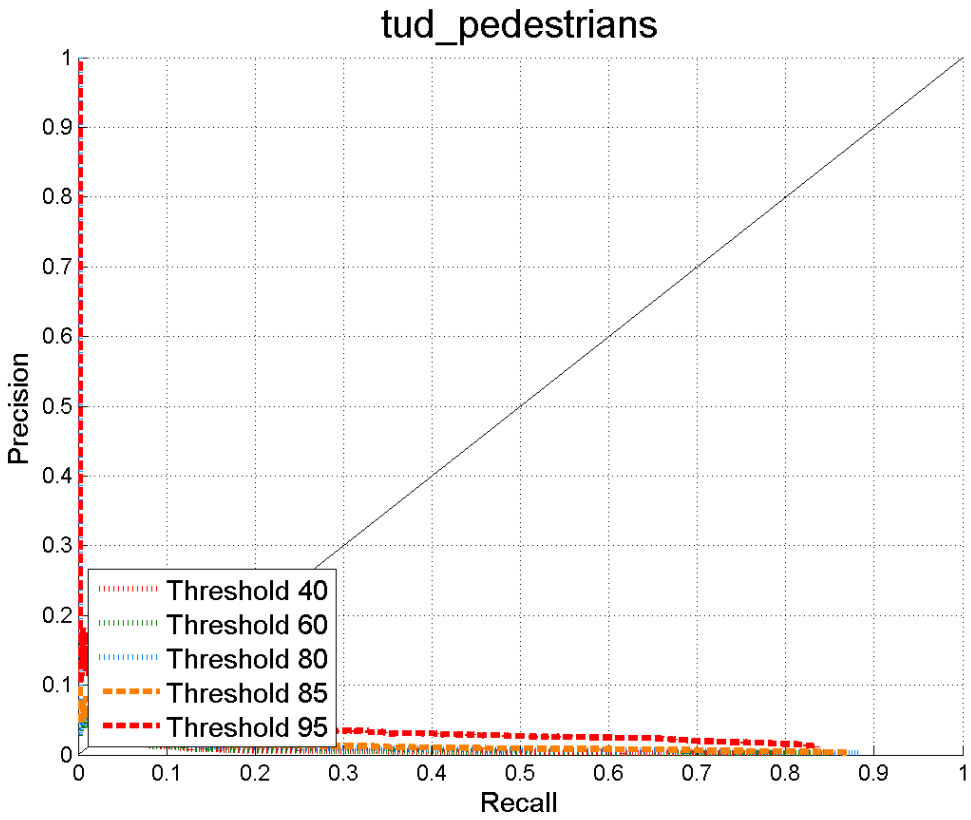

Figure 8: Different thresholds for Thresh algorithm at TUD Pedestrians

When selecting the threshold for this method is important to have in consideration some things. The first is that when using the overlap method the recall is never going to be increased. The overlap method only suppress bounding boxes and do not create new ones so the recall can only be reduced. If some bounding boxes have been created where there are no objects the Overlap method will not delete them if they are not overlaping the bounding box of a real maxima, so a low threshold is not recommended for this method also.

Different threshold were used to try to find the best one, and also to find the best possible overlap value. One of those thresholds was 0.80, in this algorithm the threshold means that the values that are under the 80% of the value of the maximum of the image are not considered. This, demonstrated that if the threshold is to low for the algorithm then the Overlap method can not solve all the problems with the precision making it be to low to be

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

considered.

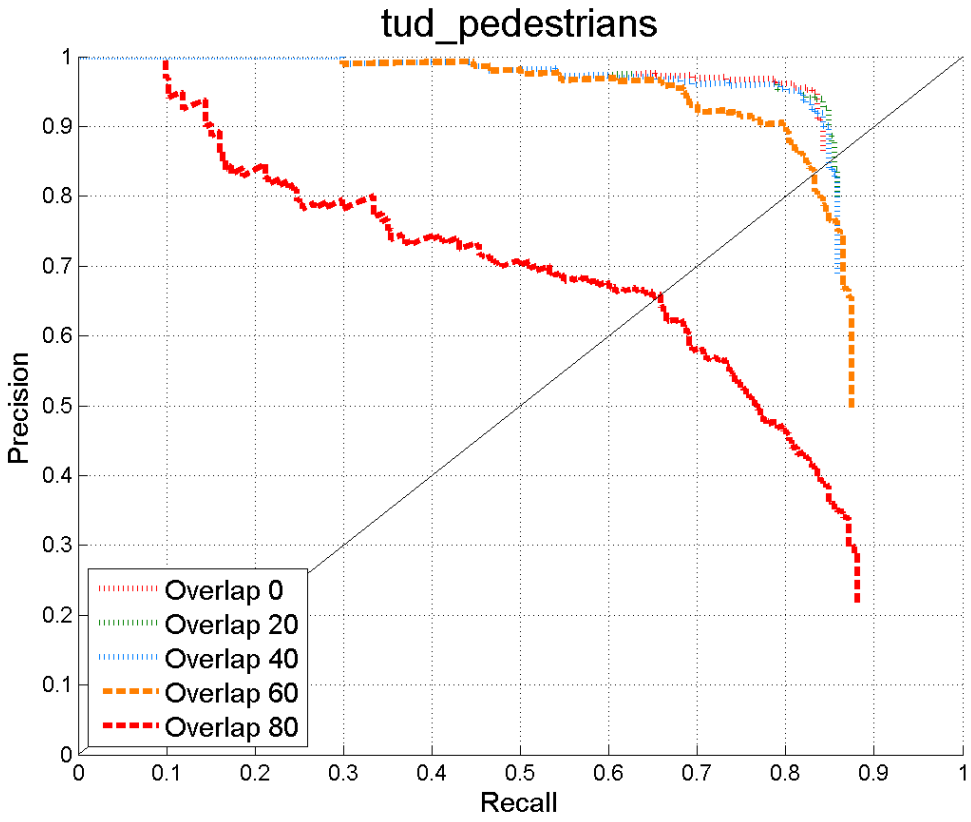

Figure 9: Different Overlaps for Thresh with Threshold of 80

Also threshold 95 was considered. In this case can be seen at the Figure 10 that the overlap method can do its job and the ones with overlap 0, 20 and 40 get both good precision and recall. Comparing the Figure 10 with the Figure 8 shows that when using the Overlap method in this case only reduce the recall slightly but increases a lot the precision.

As can be seen the best Overlap is 0, that means that the algorithms do not permit two bounding boxes to be one over the other. For this dataset this works perfectly because in it usually there are not people together so the Overlap method can suppress all the bounding boxes that are over one person without suppressing the ones of a nearly one.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

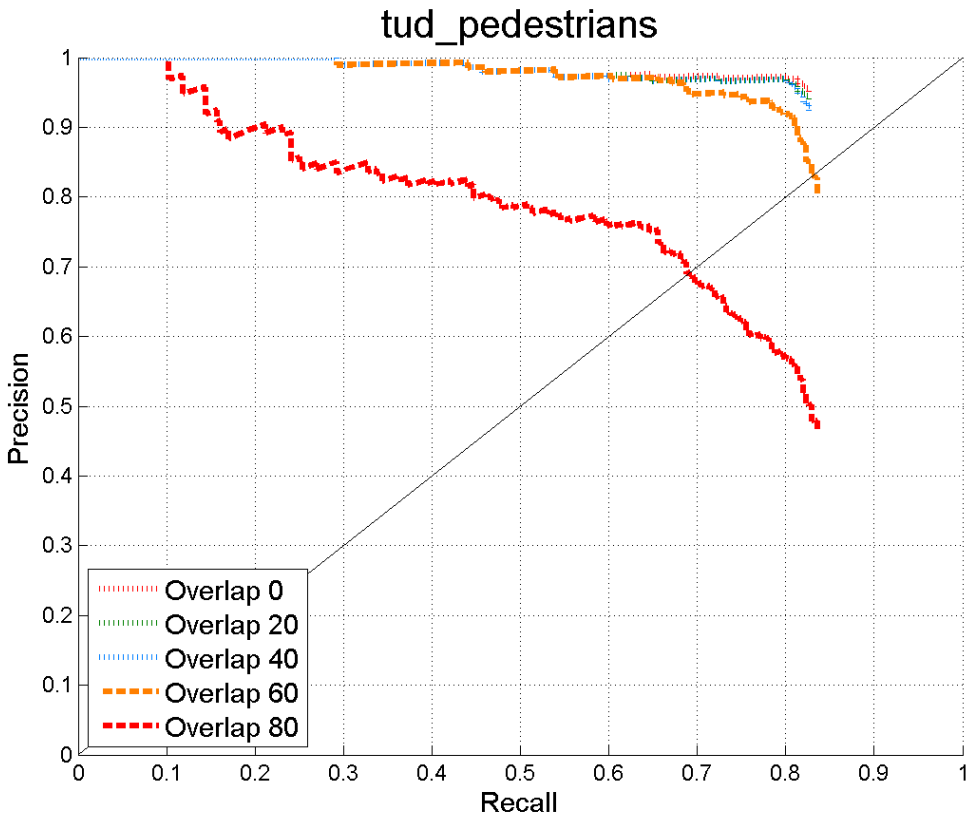

Figure 10: Different Overlaps for Thresh with Threshold of 95

4.1.4 Wenzel

As for the Thresh method in the Wenzel method before using the overlap algorithm the results have a low precision. As can be seen at the Figure 11, the maximum recall the method achieve with a 0.20 of threshold is the 90%, and of course it will be reduced once the Overlap method is used. There are some differences between the Thresh and the Wenzel response for this dataset and in general. For example the precision is a bit higher, nothing really important because the one that is in charge of the precision is mainly the Overlap algorithm.

Its interesting that this one and the Thresh method achieve almost the same recall but the Wenzel method has less computational cost. In this dataset that is because there are not much people together so when taking the local maximas in most of the cases the algorithm is not suppressing a real near maxima. This makes the Wenzel method more useful at this dataset

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

because for the same threshold, once the Overlap method is used, the output is the same for both of them but Wenzel method is faster because it works with less maximas.

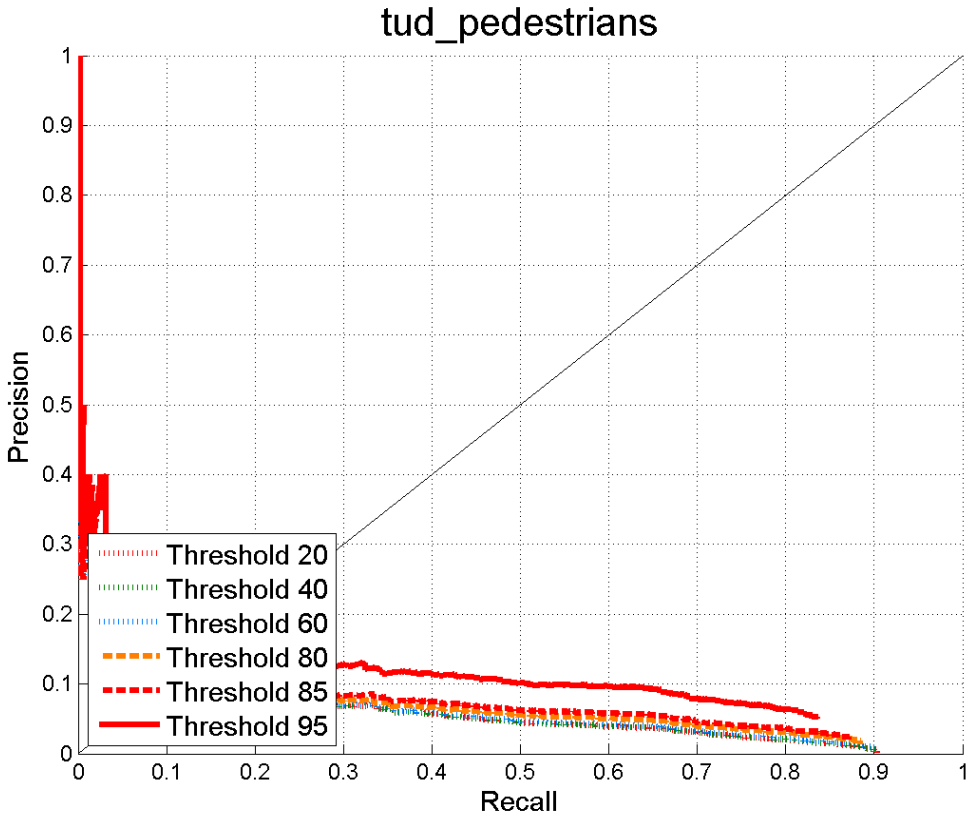

Figure 11: Different thresholds for the Wenzel method

As can be seen at the Figure 12, if it is compared with the 10 both of them look like pretty similar. Because that is not estrange to conclude that the best threshold and overlap is also in this case the same as in the Thresh case, so they are threshold 95 and overlap 0.

As explained before for this dataset the best overlap is 0 because the objects are usually alone, so nothing apart of false positives are usually suppressed even if no overlap is allowed.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

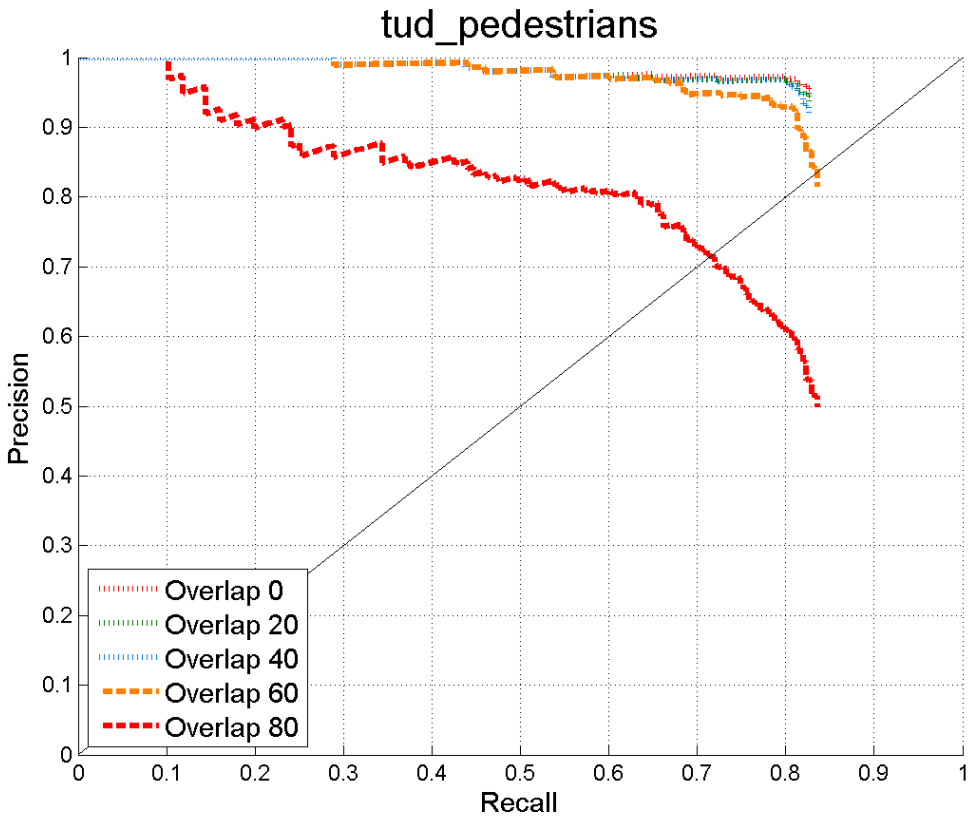

Figure 12: Different Overlaps for Wenzel with Threshold of 95

4.1.5 Islands

When the Islands method was explained also was said that this method is different in both way to work and results from the other two (Wenzel and Thresh). Now at the Figure 13 can be seen some of his differences. One of those differences is the precision of the algorithm without using the Overlap system, as can be seen its quite higher than the other two and the recall is not smaller. Also is interesting that the central thresholds (40,60,80) have a higher recall than 20 or 95, that is because with a 20% of threshold the islands are to big, and with a 95 they are to small so it means that some objects have been broken into two different islands and others have just disappeared.

As in Wenzel and Thresh in this case the threshold is a x% of the maximum value of the image. As for the other two methods this one achieves a recall near to 90% what seams to be the maximum that those methods achieve in this dataset.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

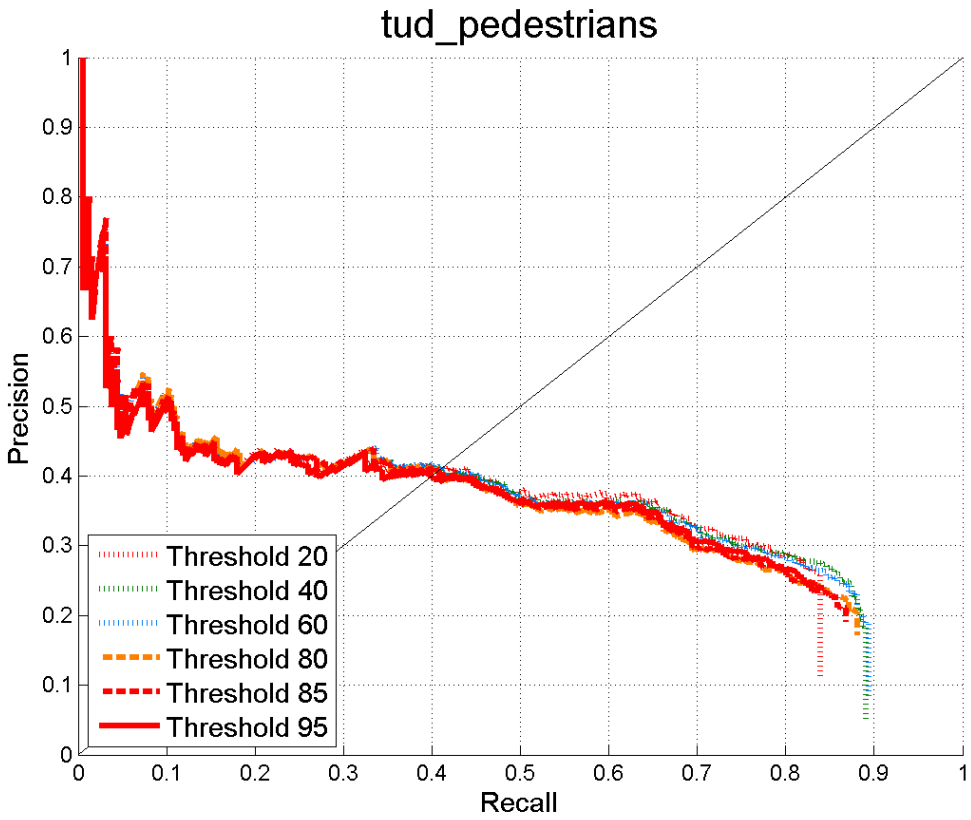

Figure 13: Different Thresholds for Islands method

It is interesting that when the threshold is high, for example 95% the islands method in this dataset starts acting as Wenzel or Thresh but with less computational cost. As said before this dataset has only between one or three people per image and they are not one over the other in most of the cases, so a high threshold makes the islands much smaller and even if before the overlap they are very different once it is used the result as seen at the Figure 14 is quite the same as in Wenzel or Thresh.

In this case again the best threshold and overlap are 95% and 0. Other thresholds were also considered such us 60% and 80% due to his high recall but the precision once used the overlap method was to low as in the case of the Thresh method so they had to be discarded for this dataset.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

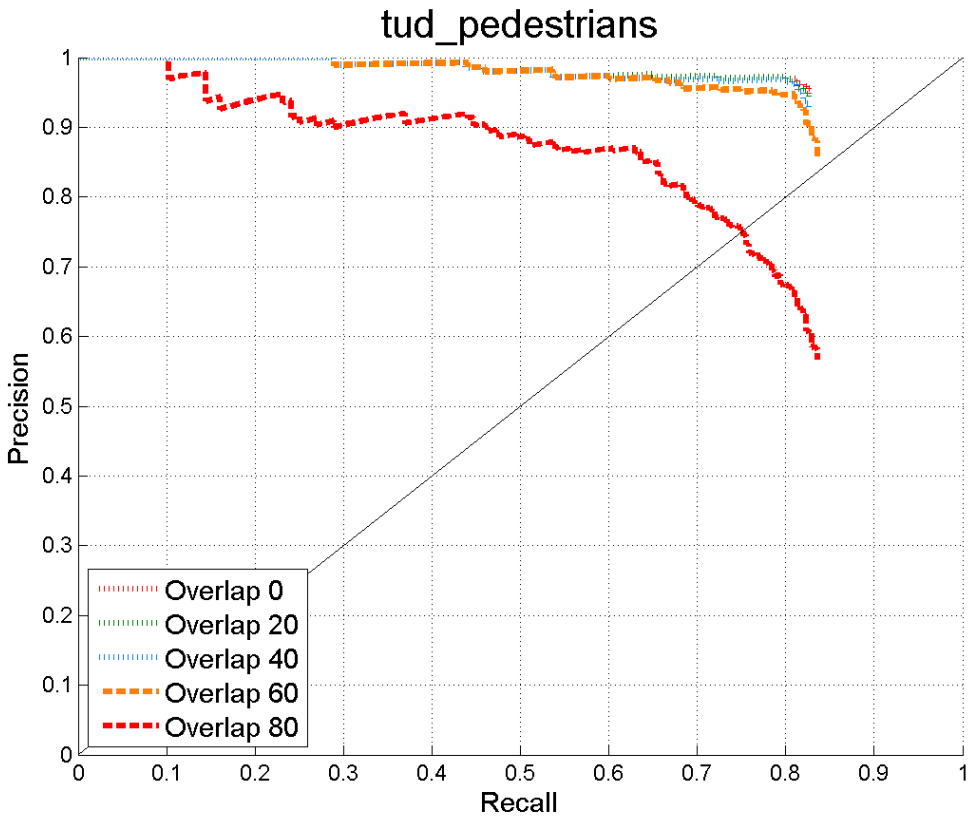

Figure 14: Different Overlaps for Islands with Threshold of 95

4.1.6 Thresh/Wenzel/Island without NMS

There are several differences between the three methods that makes their response different:

- Islands: Makes groups of pixels (islands). Lower recall for low thresholds (20%) and high thresholds (95%), and higher for the medium ones (40%,60%). Computationally faster than the others. Higher precision before using the overlap method.
- Thresh: Takes all the pixels over the threshold. High recall for low thresholds and vice versa. High computational cost. Low precision before using the overlap method.
- Wenzel: Takes the local maxima of the pixels that are over the threshold. High recall for low thresholds and vice versa. High computational

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

cost but lower than Thresh. Low precision but higher than Thresh.

Is interesting to see how similar the Thresh and Wenzel methods are and how they differ from the Islands method. While the first two methods with a threshold of 30% achieve a Recall of a 90% the Islands method only reaches 85%, even that as explained before for a threshold of 60% the recall of Islands method is near to %90 so.

Also there are differences with the precision. The Islands method achieve a higher precision even without having used the overlap method, because of the difference between it and the other methods. The Island method have in its process something similar to non-maxima suppression because the way it works. When generating the islands the algorithm already suppress part of the possible false positives while the others just take a big amount of maximas.

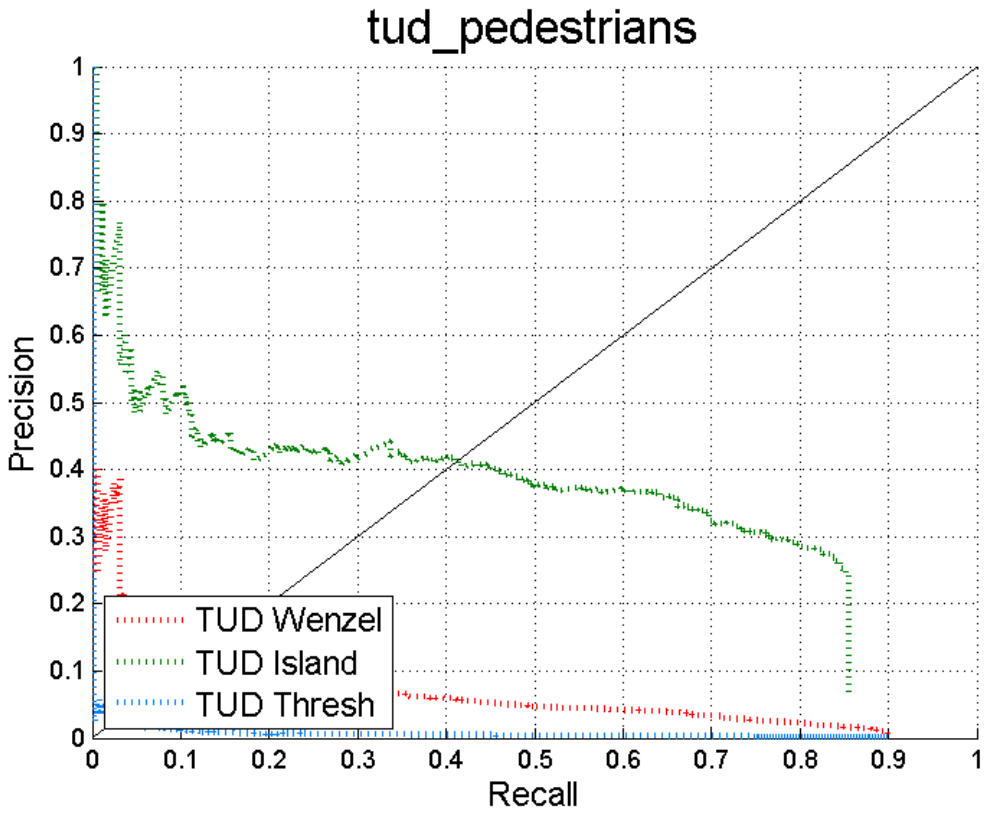

Figure 15: Thresh, Wenzel and Island methods without the Non-Maxima Suppression part.30% Threshold.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

4.1.7 Comparison

To compare the two Non-Maxima Suppression methods the best found thresholds and overlaps (in the case of Islands, Wenzel and Thresh) were used. Because Gall and the others use the threshold in a different way to be fair with all the methods the same threshold was not used for all.

When applying the NMS method to the Thresh/Wenzel/Island finding maxima algorithms the Recall drops down. It's important to say that the different methods achieve different goals. For example the Gall method has a higher recall than the other three, but the others achieve a higher precision +4-5% with only a 1-2% of less recall. It is also relevant that the three Islands, Wenzel and Thresh have the same response to that dataset at this threshold. At this point increasing the threshold of all the methods would

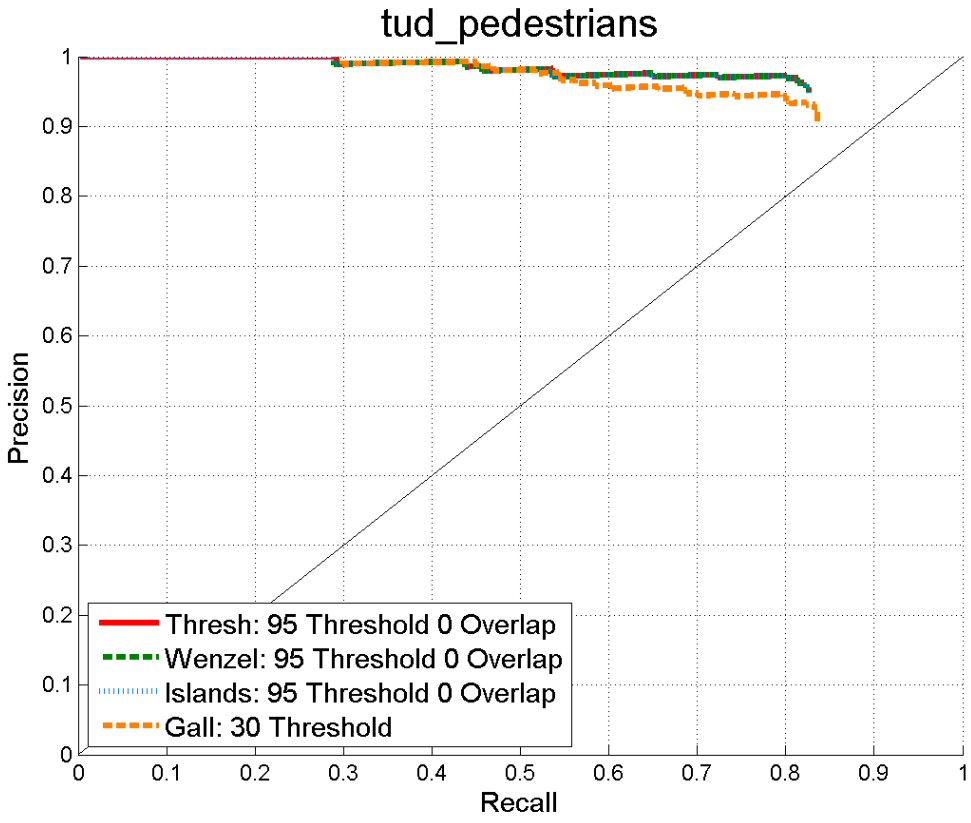

Figure 16: All the four methods compared with their best overlap and threshold.

only make the recall to decrease while the precision will be only increased a

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

little. As in all the methods the maximun recall seems to be between 85% and 90% with a very low precision, having a recall of around 83% with a high precision seems to be a good achievement.

4.2 TUD Campus

The TUD Campus dataset has some differences that makes the response of the algorithms quite different to the other dataset. In this case we also have people walking through a street, but in this case we only have one street and 71 images. The main difference in this dataset is that we have much more people walking in different directions, lots of them completely occluded or partially occluded. Having so much people together makes harder for the algorithms to select which maximas are false positives or a true positive of a nearly object. 5 scales were used in this case at the hough voting system.

Figure 17: A example of the TUD Campus dataset.

4.2.1 Gall

For this dataset the results of the Gall method were quite different to the TUD Pedestrians dataset results. As can be seen at the Figure 18 for this dataset when using Gall method the maximum recall is 83% and that is with very low thresholds, what means very low precision. In this case the 0.2

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

threshold is better than the 0.3 threshold, even if the 0.3 threshold has more precision having 100% the recall only arrives to 40%, which is extremely low.

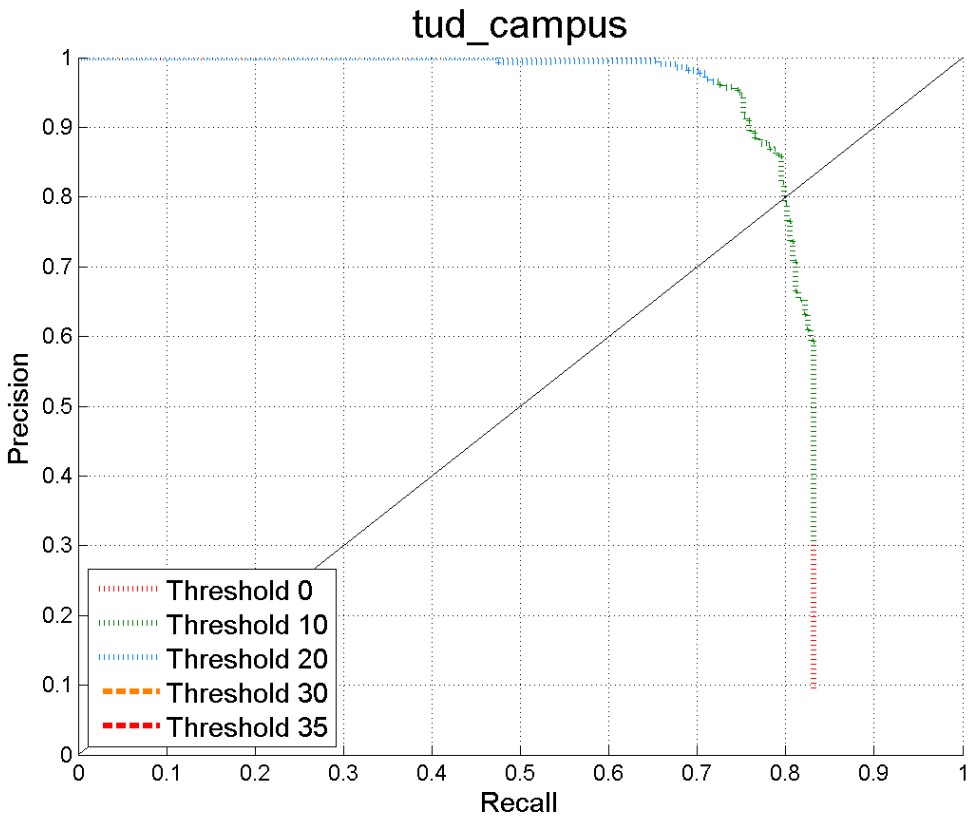

Figure 18: Gall Method for TUD Pedestrians with different thresholds

4.2.2 Gall vs Gall without interpolation

As for the other dataset the difference between using or not the interpolation is not big. The difference is only of about a 1% in both precision and recall. As seen even if the threshold is a 10% lower the recall is much smaller in both cases. This is probably because the way it works eliminating the nearly maximas. In this dataset there are to much people together and probably when eliminating some of the maximas it also eliminates other maximas that are together. When watching the images with the bounding box it's shown how it detects the people that are alone and only one or two of the groups of people.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

When the threshold is increased from 20% to 30% the precision was increased and reached a 100% but the recall got lowered to between a 50%-60%. This happens with both, interpolated and non interpolated method.

In this case the difference between both methods is smaller than the one the TUD Pedestrians. This is in part because of the higher amount of scales used. This makes easier to find the appropriated bounding box or at least don't doing it to big/small without interpolation.

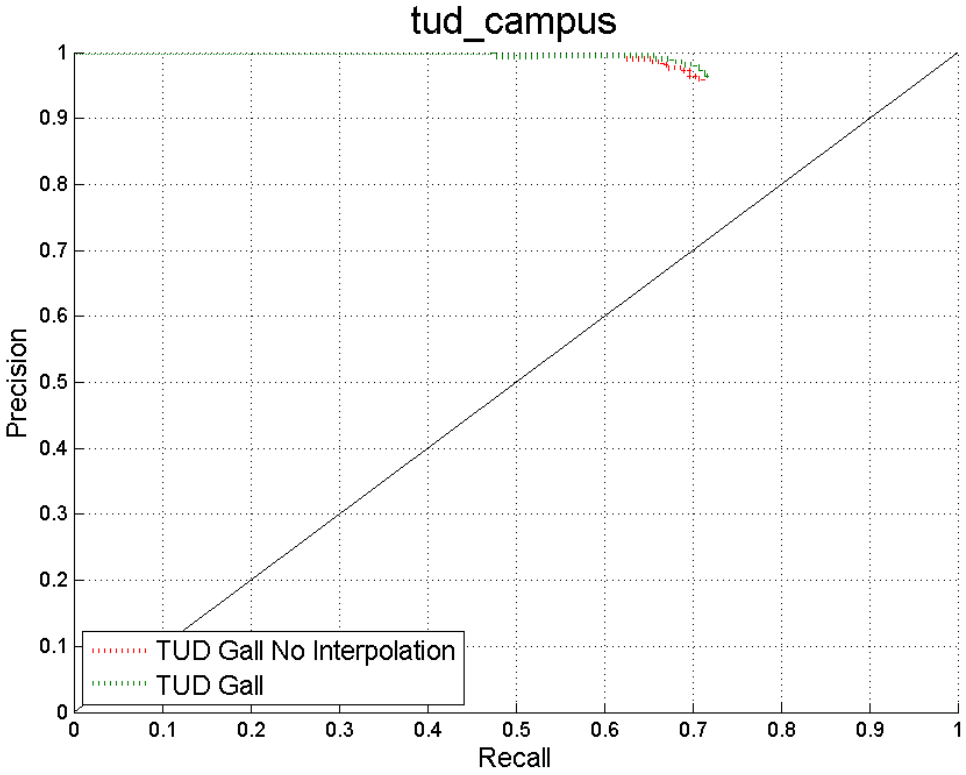

Figure 19: TUD Campus, Gall with interpolation vs Gall without interpolation. Threshold 20%

4.2.3 Thresh

The Thresh output to this dataset is the opposite of the Gall method because it reaches higher values of recall than in the other dataset. As was expected due to the previous dataset the higher recall for this dataset are reached by the 20% and 40% threshold, and all the thresholds have a low precision near to 0% because of the number of taken values.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

One of the biggest differences is that the recall for high thresholds such us 80% or 95% is extremely low. At the TUD Pedestrians dataset the recall values for all the thresholds where between 80% and 90%, but in this case the recall values for the thresholds between 80% and 95% are between 50% and 70%. That is probably because most of the people at the images are partially occluded and when creating a hough map with partial occluded people the accumulator space has less information. For example, if a person is partially occluded an only one hand, one foot and the head can be seen the hough map will be less accurate than with the two hands and the two feet, so the values of the center will have less values and with a high threshold they will not be considered.

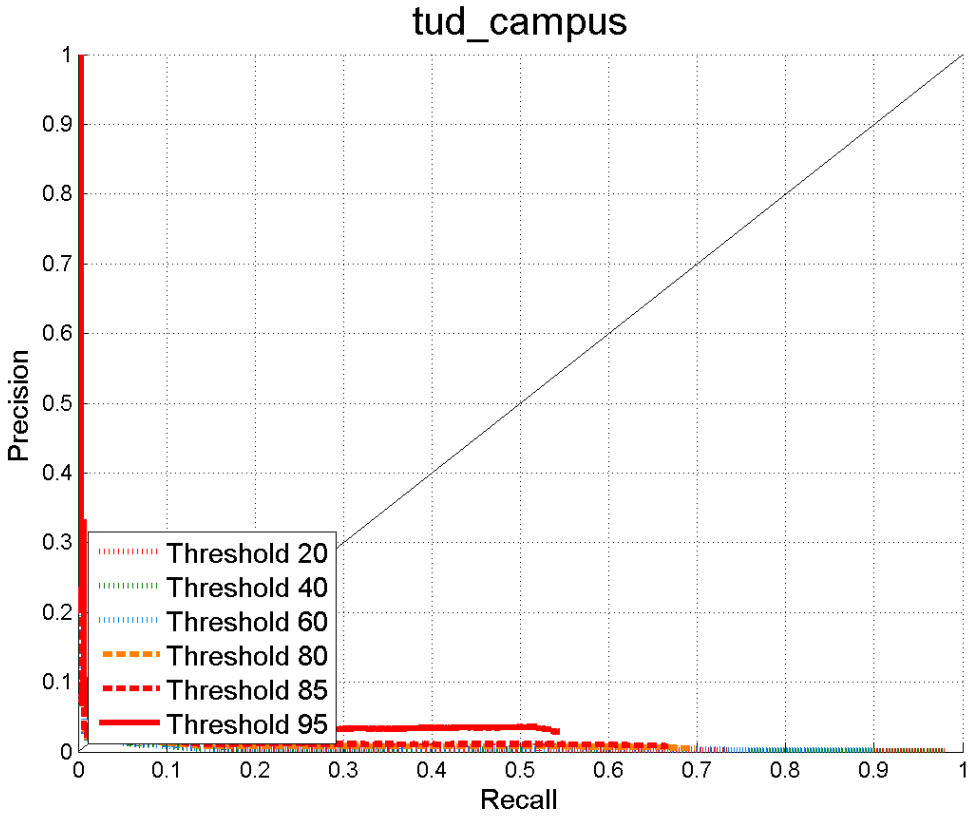

Figure 20: Thresh Method for TUD Pedestrians with different thresholds

To select the correct threshold again several different thresholds were tried with different overlaps. One of the first tries was the 40% threshold because of its recall of 98%. 40% and not 20% was chosen for a simple reason, they

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

both have the same recall but even if at the graph they seems to have the same precision its not correct. They both have a precision near to zero, but the precision of the 40% is bigger even if is not much.

Unfortunately the Overlap method could not fix the precision of the 40% threshold and was quite low for all the overlaps. However at the graph one big difference to the TUD Pedestrians dataset can be seen. At the TUD pedestrians dataset when using the Overlap method the overlap 0 was the best one and the difference between overlaps was only the precision the achieved, but in this case because of the large amount of people together is also changes the recall so its not so clear which is the best overlap.

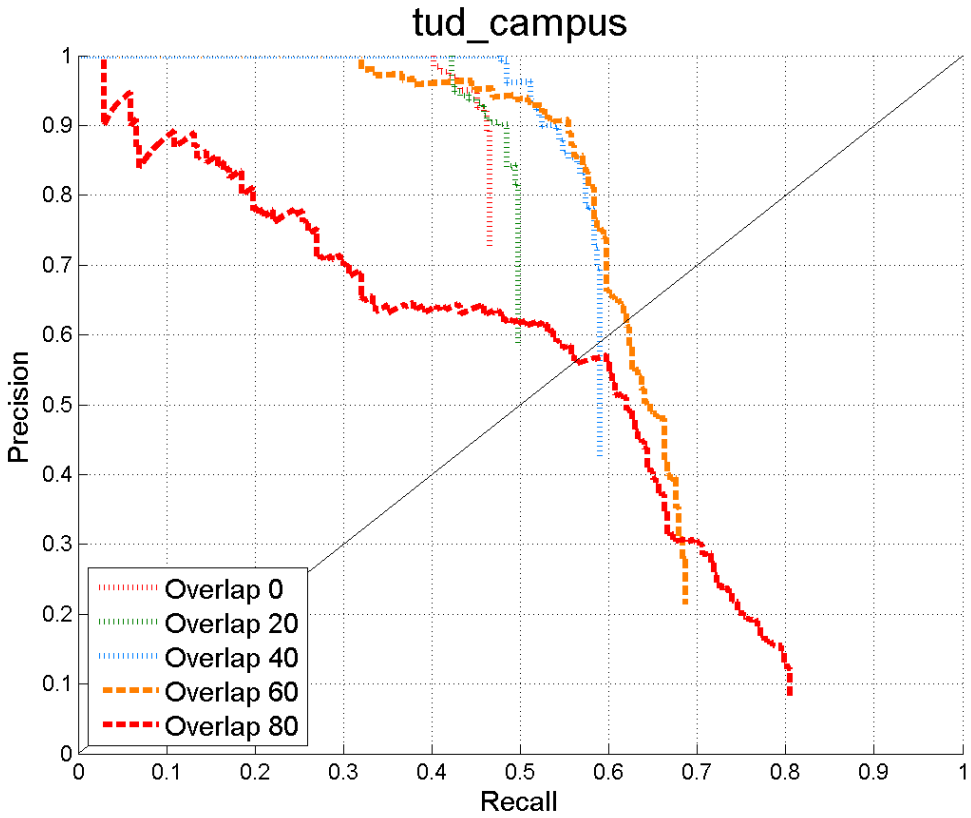

Figure 21: Different Overlaps for Thresh with Threshold of 40

Also the thresholds 60% and 80% was used and 80% was apparently the best one with an overlap of 40%. As expected the correct overlap was in this case one over the 0% overlap because the dataset. Even if this was the value considered as the best, at the graph can be seen that the recall do not reach

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

50% even if the precision is near to 100%, what means that the algorithm finds only the 50% of the objects but it does not find any false positive.

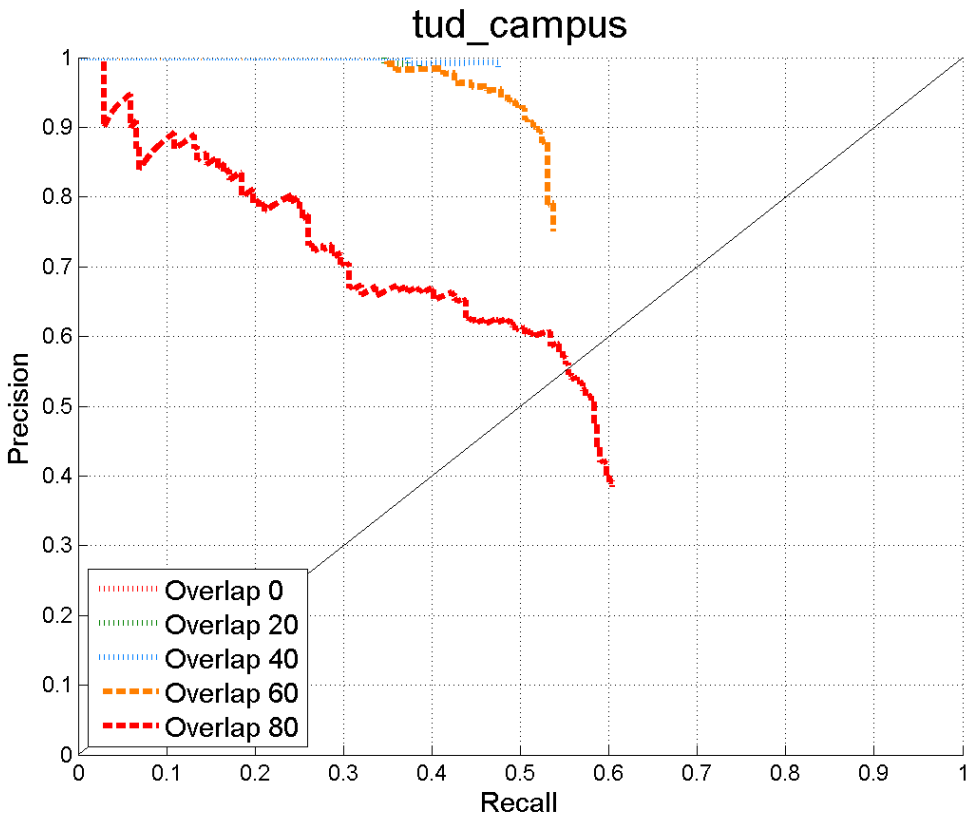

Figure 22: Different Overlaps for Thresh with Threshold of 80

4.2.4 Wenzel

For the Wenzel method at the TUD Campus dataset the differences between it and the TUD Pedestrians are quite the same as for the Thresh method. At the graph can be seen a high recall for the 20% and 40% thresholds, an a low recall for the thresholds between 80% and 95% being the difference bigger than the one found at the TUD Pedestrians dataset. The only important difference between Thresh and Wenzel is again the precision at this point even if the difference is not big.

As in the TUD Pedestrians dataset in this one the precision and recall graphs of the Wenzel and the Thresh methods are very similar, but some

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

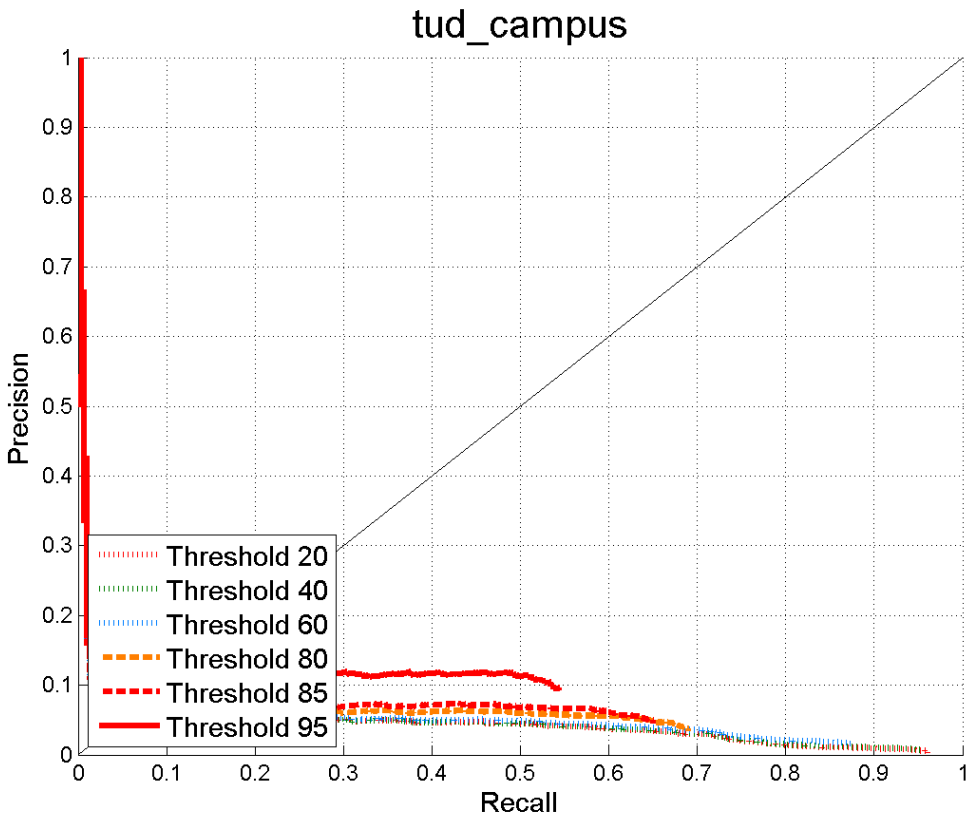

Figure 23: Wenzel Method for TUD Pedestrians with different thresholds

differences can be seen despite all the similarities. Again in this dataset the best threshold is the same for the two methods and also the best overlap, being 80% of threshold and 40% of overlap. However for the 60% and 80% overlap the Wenzel method achieves a higher recall of about 1%-2% and also a higher precision which makes it a better option if one of those have to be chosen for this dataset. Even that the recall stills being very low and does not reach the 50% of the objects.

4.2.5 Islands

The islands method as in the TUD pedestrians dataset has a different response to this dataset than the other methods. Again the central thresholds (40% and 60%) have the higher recall and the 20% the lowest one. In this case the difference is even bigger than in the other because the number of

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

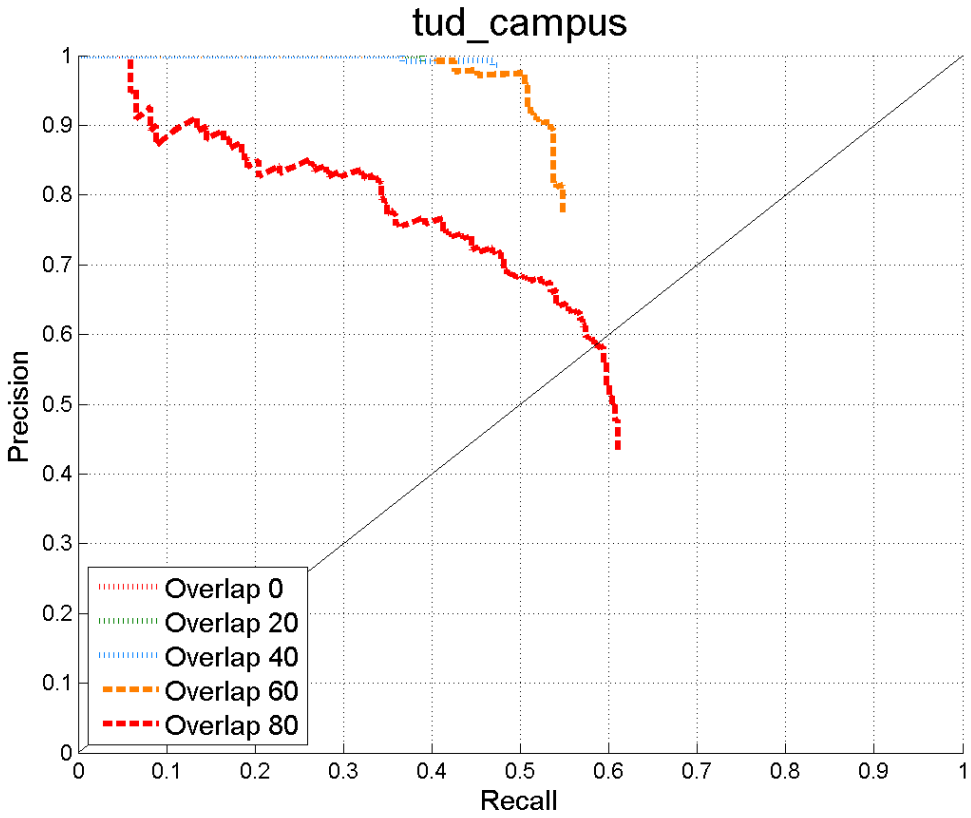

Figure 24: Different Overlaps for Wenzel with Threshold of 80

people together makes a 20% threshold to mix different people in only one island.

Also again the high thresholds have less recall than the medium ones because they do not consider some true positives and they break the island of one person in several ones.

For this dataset the best threshold is different to the ones selected for the Wenzel and Thresh methods and its 60%. In this case is difficult to decide which one is the best overlap, 20% or 40%, they both have a high precision and a low recall, 20% has around 97% of precision and 46% of recall, whereas 40% has a 92% of precision and 52% of recall, both of they are weak at recall and strong at precision.

Other thresholds that were considered did not achieve the precision in the case of the 40% threshold or the recall in the case of 80% of threshold.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

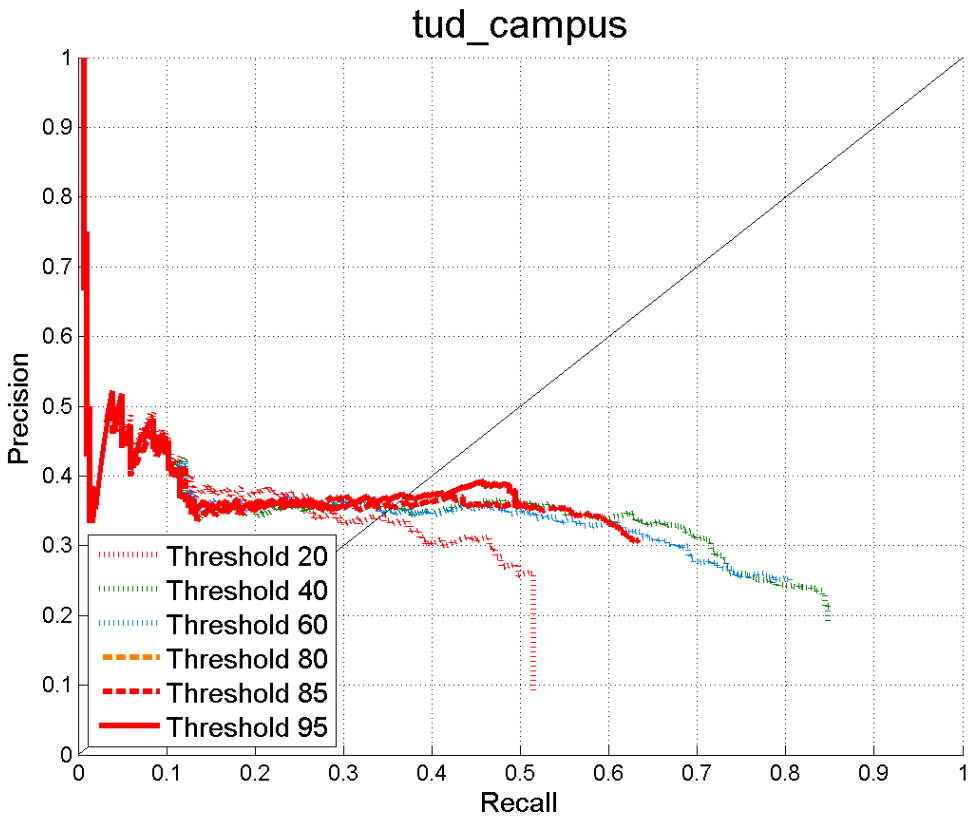

Figure 25: Islands Method for TUD Pedestrians with different thresholds

4.2.6 Thresh/Wenzel/Island without NMS

It's interesting to see how this three methods react different to different datasets and different threshold. In one hand the Island method has a very low recall to a low threshold and and a high recall to intermediate thresholds such us 40% or 60%. This is due its way to find the maximas that has already been explained. With a threshold of only the 20% and the groups of 5 people together the islands created by the algorithm are big but are the result of several people that are one over the other. As said before a higher threshold for this algorithm (not to high) makes it have a higher precision and a higher recall. This was shown at the graph of the Islands method with different thresholds.

In the other hand the Thresh and Wenzel method have again the highest recall of all the methods for a low threshold, taking in consideration that Islands method increases its recall with the threshold and Wenzel and Thresh

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

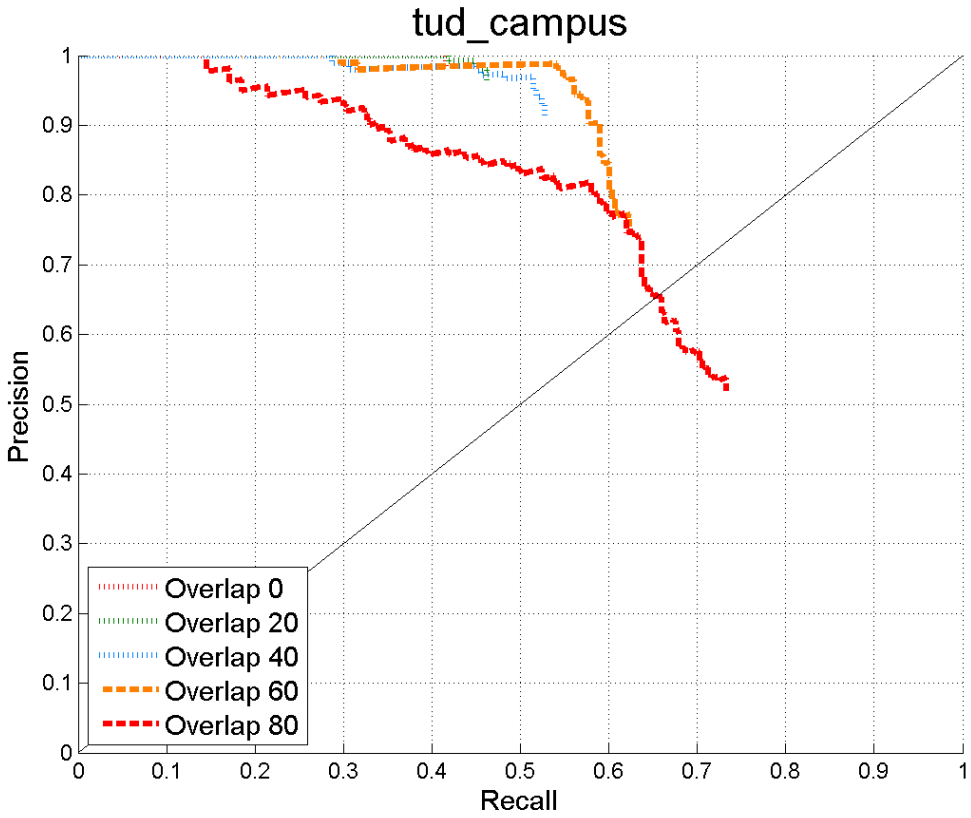

Figure 26: Different Overlaps for Islands with Threshold of 60

have it reduced, but Islands never reaches the maximun recall for Thresh and Wenzel. This change when increasing the threshold, the precision will be the around the same, but the recall would drop significantly because a higher threshold means that less weak hough mapped objects are detected. Also using a 0% threshold for the Thresh method would mean a 100% recall because all the pixels are considered maximas, whereas for the Islands would mean finding only one object per image because all the image would be one island.

4.2.7 Comparison

When compared all the four methods between them the results are quite interesting. For the best thresholds and overlaps of each of the methods, the Gall method seems to be by far the best one.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

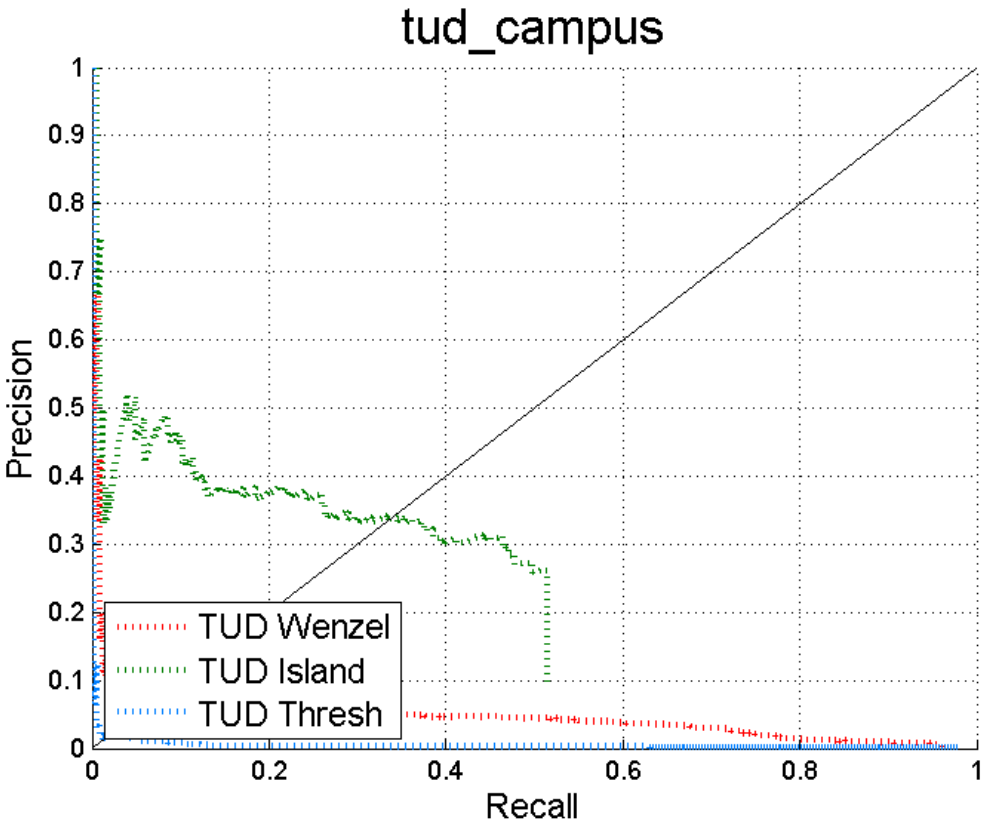

Figure 27: TUD Campus, Thresh, Wenzel and Island methods without the Non-Maxima Suppression part.20% Threshold.

All the algorithms get less recall than in the other data set but while in the TUD Pedestrians dataset the differences between them were only a few percents in this case the recall of the Gall is a 20% higher than the second highest and the precision is only a 1% or 2% lower which makes Gall the winner at this dataset.

Is also interesting that at TUD pedestrians the three, Wenzel, Thresh and Islands had the same response to the dataset, but in this case only Thresh and Wenzel share the same result and Islands get a lower precision but a higher recall, but both of them lower than Gall ones.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

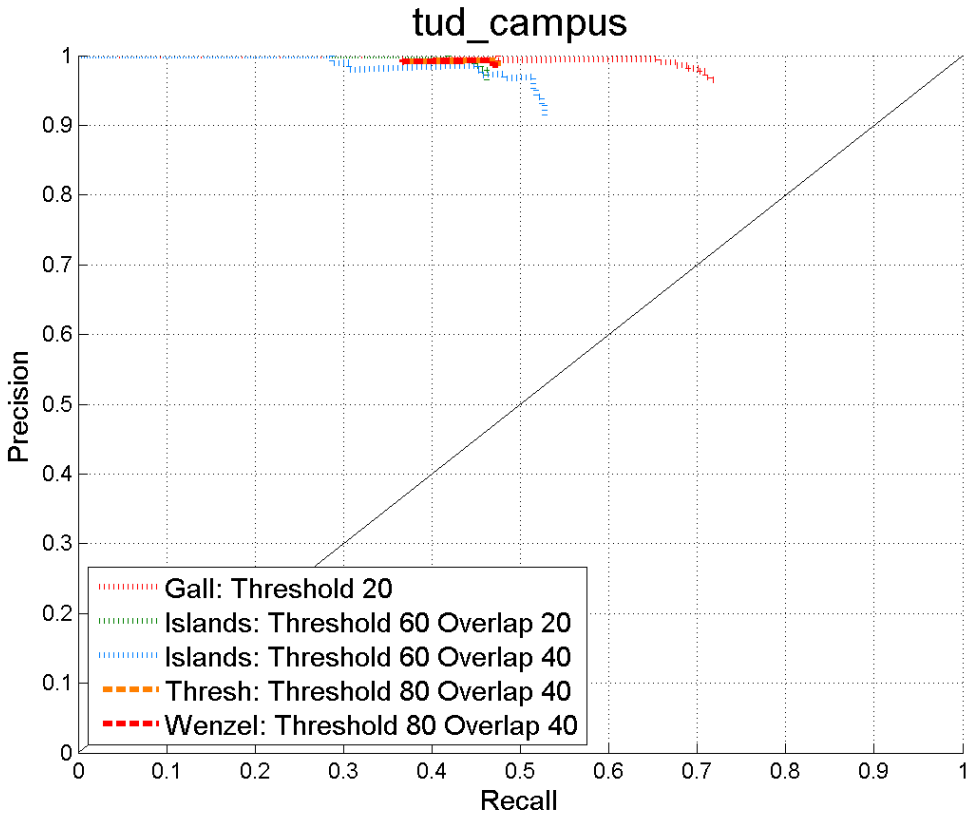

Figure 28: TUD Campus, All the four methods compared for their best thresholds and overlaps.

5 Conclusions

It is very interesting to see how for different situations such as amount of people, background, number of scales. The conclusion is that creating a algorithm for object recognition that achieves high recall and precision in all the cases and with a low computational cost is very difficult. This makes that for live applications difficult to be error free.

In this thesis the TUD Pedestrians dataset and the TUD Campus are considered like different cases, but in both of them what can be seen is people walking through a street. That makes that if someone tries to implement a method to detect people in a real street the threshold would have to be the same, also the overlap and everything and try to cover all the possible cases.

Of all the methods compared in this thesis the Gall method is the mos

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira

robust because it achieves in both datasets a good recall and a high precision, whereas the other three are very weak to the images with partially occluded people and groups of people.

Also is important to say that one of the most important steps is selecting the threshold, without it the computational cost is to high, the recall is higher but the precision is lower and as has been shown the overlap method can not fix the recall if it is to low because the algorithm finds bounding boxes in places where there are nothing but background.

With the current methods a 100% recall or a 100% precision can be achieved but not the two of them in several datasets. The challenge now is to achieve a precision and recall near 100% for different datasets to cover as much cases of the life as possible with a low computational cost.

References

[1] J.Gall and V.Lempitsky. Class-specific hough forests for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[2] D.McAllester and D.Ramanan. A discriminatevely trained, multiscale, deformable part model. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[3] B. Leibe, A. Leonardis, and B.Schiele. Robust object detection with interleaved categorization and segmentation. In IJCV Special Issue on Learning for Recognition and Recognition for Learning, 2008.

upna Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa
Todos los derechos reservados
Eskubide guztiak erresalbatu dira