Manhattan-Pyramid Distance: a Solution to an Anomaly in Pyramid Matching by Minimization

Aneesh Chauhan and Luís Seabra Lopes

IEETA/DETI, Universidade de Aveiro, Aveiro, 3810-193, Portugal
{aneesh.chauhan, lsl}@ua.pt

Abstract

In the field of computer vision, pyramid matching by minimization has gained increasing popularity. This paper points out and discusses an inherent anomaly in pyramid matching by minimization that can affect the performance of classification approaches based on this type of matching. As a solution, a new multi-resolution measure, called Manhattan-Pyramid Distance ($MPD$), is proposed. Systematic evaluations are carried out at the task of instance-based object classification on four object image datasets. Results show that $MPD$ improves object classification performance with respect to a standard approach based on pyramid matching by minimization.

1. Introduction

Matching by minimization is at the core of various kernel functions based on multi-resolution analysis, namely the Pyramid Match Kernel ($PMK$) \cite{4, 5} and its variants (e.g. spatial pyramid matching \cite{2, 7}). These approaches have gained increasing attention, especially in the field of visual categorization \cite{10, 11, 12}. $PMK$ was designed to enable the application of kernel-based learning methods to domains where objects are represented by unordered and variable-sized sets of features, such as sets of local features in computer vision.

In this matching approach, each feature set is mapped to a histogram pyramid, i.e. a multi-resolution histogram preserving the individual features distinctness at the base level. Then, the histogram pyramids are matched using a weighted histogram intersection at each resolution level. By definition, histogram intersection sums the minimum values in corresponding bins, thus, we come to describe the overall approach as pyramid matching by minimization. This more general term is useful, especially because the approach is applicable to pyramids in which pyramid levels are (multi-)dimensional vectors, but not necessarily histograms.

Pyramid matching can capture similarities at multiple resolutions, which is an advantage with respect to other simpler metrics that compute (dis)similarity at a single resolution. One problem with pyramid matching by minimization, as done in $PMK$, is precisely the use of minimization. Similarity is directly related with proximity (or inversely related with distance) \cite{6}, but this relation is not captured when matching by minimization. In this paper, we will elaborate on the anomaly imposed by ‘matching by minimization’ and propose a new multi-resolution distance measure called Manhattan-Pyramid Distance ($MPD$) which resolves the anomaly.

The new measure is evaluated at the task of instance-based object classification on four object image datasets. In this context, $MPD$ is converted to a similarity for more direct comparison with $PMK$.

The rest of the paper is organized as follows: Section 2 briefly presents the standard algorithm for pyramid matching by minimization and discusses the mentioned anomaly. Section 3 presents the new multi-resolution measure based on Manhattan distance. Section 4 describes the classification scenario in which the similarity measures will be evaluated. Section 5 describes the performed experiments as well as the obtained results.

2. Pyramid matching by minimization

2.1. Basic algorithm

The basic algorithm for pyramid matching by minimization, which forms the core of $PMK$ \cite{4, 5}, takes as input two feature vectors $\bar{x}$ and $\bar{y}$ representing two objects to be compared. These feature vectors can be an abstraction of more complex/raw representations. For instance, in the context of $PMK$, the matched feature
vectors are histograms computed from unordered and variable-sized points sets. Each point in these point sets can itself be represented as a feature vector in some other feature space. The matching algorithm then combines similarity scores between \( \bar{x} \) and \( \bar{y} \) at increasingly coarse resolutions. More specifically, the final pyramid match score is computed as a weighted sum of the partial scores, where coarser resolutions are given less weight:

\[
\hat{P}_\Delta(\bar{x}, \bar{y}) = \sum_{i=0}^{L-1} w_i N_i(\bar{x}, \bar{y})
\]

where \( L \) is the number of pyramid layers, \( w_i = 1/2^i \) is the weight of layer \( i \) and \( N_i \) measures the additional matching at layer \( i \), as given by:

\[
N_i(\bar{x}, \bar{y}) = I(F_i(\bar{x}), F_i(\bar{y})) - I(F_{i-1}(\bar{x}), F_{i-1}(\bar{y}))
\]

where \( F_i(\bar{x}) \) is the feature representation of object \( \bar{x} \) at layer \( i \) and \( I() \) is an intersection function that measures the overlap of two objects as follows:

\[
I(A, B) = \sum_{j=1}^{r} \text{min}(A_j, B_j)
\]

where \( A \) and \( B \) are feature vectors of size \( r \), and \( A_i \) is the value of \( i^{th} \) element of \( A \).

Although presented here for single-dimension vectors, the extension of this algorithm to the multi-dimensional case is straightforward. Moreover, although originally designed for working with histograms, this matching algorithm is also applicable to other types of feature vectors [11].

2.2. The anomaly in matching by minimization

Similarity is directly related with proximity (or inversely related with distance) [6], but this relation is not captured when matching by minimization. For example, Euclidean similarity (defined as the inverse of Euclidean distance) is more robust when an object, with low feature values, is compared with other objects, with higher, but clearly different feature values. An extreme situation, for a feature space with a single feature, is illustrated in Fig. 1: \( \bar{o} = \{1\} \), \( \bar{p} = \{2\} \) and \( \bar{q} = \{10\} \).

With Euclidean similarity, we get \( ES(\bar{o}, \bar{p}) = 1 \), \( ES(\bar{o}, \bar{q}) = 1/9 \), \( ES(\bar{p}, \bar{q}) = 1/8 \), which correctly describes the situation, i.e. \( \bar{o} \) and \( \bar{p} \) would still be more similar to each other than to \( \bar{q} \). However, both the un-normalized pyramid match, \( \hat{P}_\Delta \), and the normalized one, \( P_\Delta \), would lead to the conclusion that \( \bar{p} \) and \( \bar{q} \) were more similar to each other than to \( \bar{o} \).

2.3. Normalized Pyramid Match

In the context of \( MK \), to avoid favoring larger input sets (which translate into histograms with larger values), Grauman and Darrell [4, 5] proposed to normalize the pyramid match score by the product of the self similarities of the input histograms:

\[
P_\Delta(\bar{x}, \bar{y}) = \frac{\hat{P}_\Delta(\bar{x}, \bar{y})}{\sqrt{\hat{P}_\Delta(\bar{x}, \bar{x}) \cdot \hat{P}_\Delta(\bar{y}, \bar{y})}}
\]

This also resolves the extreme case described above. However, in borderline cases, the anomaly persists even after normalization. As an example, if the feature of the second object in the previous case, \( \bar{p} \), was 4, rather than 2, \( ES \) would still correctly describe the situation, i.e. \( \bar{o} \) and \( \bar{p} \) would still be more similar to each other than to \( \bar{q} \). However, both the un-normalized pyramid match, \( \hat{P}_\Delta \), and the normalized one, \( P_\Delta \), would lead to the conclusion that \( \bar{p} \) and \( \bar{q} \) were more similar to each other than to \( \bar{o} \).

3. Manhattan Pyramid Distance

To combine the advantages of distance information with multi-resolution analysis in the style of \( MK \), we designed a new measure, the Manhattan-Pyramid Distance (\( MPD \)), which is also computed at multiple
resolution scales. MPD is computed by subtracting a weighted sum of distance reductions, obtained by successively lowering the resolution, from a distance computed at the base resolution level. In each step of the algorithm, resolution is divided by 2. All distances are measured using the well known Manhattan Distance (MD) measure, defined as follows:

\[ MD(\bar{x}, \bar{y}) = \sum_{i=1}^{r} |\bar{x}_i - \bar{y}_i| \]  

(5)

where \( \bar{x} \) and \( \bar{y} \) are feature vectors representing two objects and \( r \) is the number of features.

The distance reduction or discount at resolution level \( i \), \( R_i(\bar{x}, \bar{y}) \), is given by the difference between the Manhattan distances at resolution \( i - 1 \) and \( i \):

\[ R_i(\bar{x}, \bar{y}) = MD(F_{i-1}(\bar{x}), F_{i-1}(\bar{y})) - MD(F_i(\bar{x}), F_i(\bar{y})) \]  

(6)

where \( F_i(\bar{x}) \) is the feature vector representing object \( \bar{x} \) at resolution level \( i \). The final discounted distance is given by:

\[ MPD(\bar{x}, \bar{y}) = MD(\bar{x}, \bar{y}) - \sum_{i=1}^{L-1} w_i R_i(\bar{x}, \bar{y}) \]  

(7)

where \( L \) is the number of resolution levels and \( w_i = 1/2^i \) is the weight of the distance reduction at level \( i \) in the final discount.

To convert a distance metric to a similarity measure, multiple approaches can be used. Two classical approaches are [1, 6]: inverse of distance (1/D); and exponential similarity \( \exp(-\alpha D^2) \), where \( \alpha \) and \( \beta \) are the constants that respectively give the slope and decay of the exponential function.

In the experiments presented in this paper, for more direct comparison with other similarity measures, we define the Manhattan-Pyramid similarity measure, MPS, as the inverse of MPD:

\[ MPS(\bar{x}, \bar{y}) = 1/MPD(\bar{x}, \bar{y}) \]  

(8)

4. Classification scenario

The similarity measures described above will be evaluated over multiple feature spaces and decision rules at the task of visual object classification.

In this paper, four different feature spaces, all corresponding to different types of shape signatures, are used for evaluating the presented similarity measures. These feature spaces are extracted from the edges of an object by segmenting the smallest circle enclosing the object and centered in its geometric center. Two basic types of segmentations are used: slices and concentric layers (see Fig. 2). Current implementation uses \( NS = 40 \) slices and \( NL = 40 \) layers.

The feature spaces, borrowed from our previous work [11], are the following:

- **Shape layers histogram (SLH).** The histogram contains, for each layer, the percentage of edge pixels with respect to the total number of edge pixels of the object.

- **Shape slices histogram (SSH).** The histogram contains, for each slice, the percentage of edge pixels in that slice with respect to the total number of edge pixels of the object.

- **Shape slices normalized radii averages (SSNRA).** For each slice, \( i \), the average radius of all edge pixels in that slice, \( R_i \), is computed. In this feature space, an object is represented by a vector \( \vec{r} = r_1 ... r_{NS} \), where \( r_i = R_i/R \) and \( R \) is the average of all \( R_i \).

- **Shape slices normalized radii standard deviations (SSNRSD).** For each slice, \( i \), the radius standard deviation of all pixels in that slice, \( S_i \), is computed. In this feature space, an object is represented by a vector \( s = s_1 ... s_{NS} \), where \( s_i = S_i/R \) and \( R \) is the average radius, as above.

All the feature spaces are scale-invariant, but only the first feature space is also rotation-invariant. For slice-based feature spaces, similarity between any two elements is computed as the maximum similarity between the respective feature vectors as they are circularly rotated relative to each other.

In the classification scenario designed for evaluation, object categories are represented by sets of known instances and a classifier is setup by selecting a feature space, a measure of similarity between objects and a decision rule. The considered similarity measures are MPS (eq. 8), \( P_\Delta \) (eq. 4) and Euclidean similarity,
Two decision rules are considered, namely nearest neighbor (NN) and average (AVG). For classifiers based on the AVG rule, the membership of the target object \( y \) to an object category \( O \) is measured as:

\[
M(y, O) = \text{average}_{x \in O} S(y, x)
\]

where \( S(y, x) \) is the similarity between objects \( x \) and \( y \). For the classifiers based on the NN rule, the membership is given by:

\[
M(y, O) = \max_{x \in O} S(y, x)
\]

The target object \( y \) is labeled with the category \( O \) for which the highest membership is obtained.

5. Experimental Evaluation

5.1. Datasets

The experiments were carried out on four datasets: LANGG68 [11] which contains 7350 color images of 68 object categories; COIL-100 [9] which contains 7200 color images of 100 different objects; ETH80 [8] consisting of 3526 images of 100 different categories; and ALOI-1000 [3], the largest dataset, containing 110,250 images of 1000 objects. For every dataset, the object images were captured using a fixed camera against a black background and each image contains a single object.

The main difference between these datasets is that some of the object categories in LANGG68 and all the categories in ETH80 contain more than one object, whereas each category in COIL-100 and ALOI-1000 strictly contain images of the same object. In LANGG68, all images of an object were captured with the object in the same pose. In the remaining datasets, objects appear in different poses in different images. COIL-100 dataset is the only evenly distributed dataset, that is, each category has exactly 72 images, each corresponding to a different pose of the object. Images in ALOI-1000 are clearly more noisy than those in the other datasets.

5.2. Results

For each basic classifier configuration, one 10-fold cross-validation experiment was carried out on every dataset. Each configuration includes the choice of a feature space (SLH, SSH, SSNRA and SSNRSID), similarity measure (ES, \( P_\Delta \) and MPS) and the classification rule (NN, AVG). Table 2 shows average accuracies obtained with each shape signature, similarity measure and decision rule. The different classifier configurations, 24 in total, and respective results are listed in Table 1.

It can be seen from this table that some of the best results are obtained with the SSNRA and SSH feature spaces and NN decision rule. More importantly, MPS performed clearly above \( P_\Delta \) when the AVG decision rule was used, but only marginally above \( P_\Delta \) when the NN decision rule was used. This suggests that some learning and classification strategies can suffer more from the matching by minimization anomaly than others.

In most cases (27 out of 32), the configurations with MPS led to the highest classification accuracy (in contrast, \( P_\Delta \) led to the best result in only 8 cases, and ES in only 2 cases). Results from Table 2 also clearly show that, over all datasets, MPS similarity measure, proposed in this paper, outperforms both ES and \( P_\Delta \). Also note, \( P_\Delta \) performed the poorest in terms of average accuracy for all datasets. We believe that the poor performance of \( P_\Delta \) is caused, to a large extent, by the anomaly pointed out earlier.

| Table 1. Average performance of the different shape signatures, similarity measures and decision rules |
|-----------------------------------------------|----------|---------|----------|
| Average acc. (%)                             | LANGG68  | COIL-100 | ETH80    | ALOI-1000 |
| Similarity measure                           |          |          |          |           |
| ES                                            | 61.7     | 49.9     | 52.5     | 42.4      |
| MPS                                           | 64.4     | 53.2     | 55.6     | 44.5      |
| \( P_\Delta \)                                | 59.9     | 47.7     | 52.4     | 38.1      |
| Feature space                                 |          |          |          |           |
| SLH                                           | 60.7     | 50.9     | 55.1     | 36.2      |
| SSH                                           | 58.8     | 53.9     | 55.1     | 46.6      |
| SSNRA                                         | 72.6     | 52.1     | 55.7     | 45.9      |
| SSNRSID                                       | 55.8     | 44.3     | 48.1     | 38.1      |
| Decision rule                                 |          |          |          |           |
| AVG                                           | 50.3     | 29.0     | 41.5     | 20.6      |
| NN                                            | 73.7     | 71.5     | 65.5     | 62.8      |

6. Conclusions

A new multi-resolution measure of distance, Manhattan-Pyramid Distance \( MPD \), was proposed. This measure was designed to rectify the anomaly in pyramid matching by minimization. Systematic experiments were conducted to evaluate and compare the performance of \( MPD \) against \( P_\Delta \) and ES, at the task of object classification on four object image datasets. These measures were evaluated over four different feature spaces and two decision rules. From these evaluations, two main conclusions can be derived: overall,
MPD outperforms both $P_\Delta$ and ES; and the anomaly in matching by minimization affects some learning and classification strategies more than others.

7. ACKNOWLEDGMENT

This work was partially funded by COMPETE/FEDER and Portuguese science foundation (FCT) in the context of the project FCOMP-01-0124-FEDER-022682.

References


<table>
<thead>
<tr>
<th>Feature space</th>
<th>Similarity measure</th>
<th>Classifier type</th>
<th>Avg. acc. ± stdev (%)</th>
<th>LANGG68</th>
<th>COIL-100</th>
<th>ETH80</th>
<th>ALOI-100</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLH</td>
<td>ES</td>
<td>AVG</td>
<td>48.5 ± 1.42</td>
<td>32.8 ± 1.01</td>
<td>41.5 ± 2.96</td>
<td>21.2 ± 0.27</td>
<td></td>
</tr>
<tr>
<td>SLH</td>
<td>MPS</td>
<td>AVG</td>
<td>54.2 ± 1.43</td>
<td>38.5 ± 1.62</td>
<td>49.3 ± 2.67</td>
<td>22.9 ± 0.35</td>
<td></td>
</tr>
<tr>
<td>SLH</td>
<td>$P_\Delta$</td>
<td>AVG</td>
<td>45.9 ± 1.06</td>
<td>27.9 ± 1.28</td>
<td>41.6 ± 0.71</td>
<td>12.2 ± 0.19</td>
<td></td>
</tr>
<tr>
<td>SLH</td>
<td>ES</td>
<td>NN</td>
<td>70.3 ± 1.68</td>
<td>65.6 ± 1.21</td>
<td>64.6 ± 1.55</td>
<td>52.7 ± 0.38</td>
<td></td>
</tr>
<tr>
<td>SLH</td>
<td>MPS</td>
<td>NN</td>
<td>72.5 ± 1.67</td>
<td>70.0 ± 1.04</td>
<td>67.3 ± 3.02</td>
<td>54.1 ± 0.37</td>
<td></td>
</tr>
<tr>
<td>SLH</td>
<td>$P_\Delta$</td>
<td>NN</td>
<td>72.5 ± 1.39</td>
<td>69.8 ± 1.54</td>
<td>66.3 ± 2.29</td>
<td>54.1 ± 0.35</td>
<td></td>
</tr>
<tr>
<td>SSH</td>
<td>ES</td>
<td>AVG</td>
<td>48.2 ± 1.62</td>
<td>30.5 ± 0.91</td>
<td>45.7 ± 2.83</td>
<td>24.7 ± 0.35</td>
<td></td>
</tr>
<tr>
<td>SSH</td>
<td>MPS</td>
<td>AVG</td>
<td>50.8 ± 1.32</td>
<td>35.5 ± 1.32</td>
<td>37.3 ± 1.45</td>
<td>27.5 ± 0.33</td>
<td></td>
</tr>
<tr>
<td>SSH</td>
<td>$P_\Delta$</td>
<td>AVG</td>
<td>43.6 ± 1.16</td>
<td>24.6 ± 0.89</td>
<td>39.5 ± 1.86</td>
<td>13.9 ± 0.22</td>
<td></td>
</tr>
<tr>
<td>SSH</td>
<td>ES</td>
<td>NN</td>
<td>67.9 ± 1.48</td>
<td>75.0 ± 1.41</td>
<td>62.7 ± 2.46</td>
<td>69.4 ± 0.40</td>
<td></td>
</tr>
<tr>
<td>SSH</td>
<td>MPS</td>
<td>NN</td>
<td>71.6 ± 0.95</td>
<td>78.9 ± 1.96</td>
<td>67.5 ± 3.44</td>
<td>72.0 ± 0.35</td>
<td></td>
</tr>
<tr>
<td>SSH</td>
<td>$P_\Delta$</td>
<td>NN</td>
<td>70.9 ± 1.28</td>
<td>78.7 ± 1.23</td>
<td>67.7 ± 1.66</td>
<td>72.1 ± 0.34</td>
<td></td>
</tr>
<tr>
<td>SSNRA</td>
<td>ES</td>
<td>AVG</td>
<td>65.1 ± 1.17</td>
<td>31.2 ± 0.94</td>
<td>39.2 ± 1.48</td>
<td>25.5 ± 0.27</td>
<td></td>
</tr>
<tr>
<td>SSNRA</td>
<td>MPS</td>
<td>AVG</td>
<td>66.3 ± 1.52</td>
<td>31.7 ± 1.41</td>
<td>41.5 ± 1.78</td>
<td>27.6 ± 0.44</td>
<td></td>
</tr>
<tr>
<td>SSNRA</td>
<td>$P_\Delta$</td>
<td>AVG</td>
<td>56.2 ± 1.38</td>
<td>20.3 ± 0.84</td>
<td>38.4 ± 1.63</td>
<td>12.4 ± 0.23</td>
<td></td>
</tr>
<tr>
<td>SSNRSN</td>
<td>ES</td>
<td>AVG</td>
<td>82.7 ± 1.2</td>
<td>75.9 ± 1.3</td>
<td>71.8 ± 2.50</td>
<td>68.5 ± 0.33</td>
<td></td>
</tr>
<tr>
<td>SSNRSN</td>
<td>MPS</td>
<td>AVG</td>
<td>82.7 ± 0.83</td>
<td>76.8 ± 1.18</td>
<td>71.4 ± 2.04</td>
<td>70.8 ± 0.41</td>
<td></td>
</tr>
<tr>
<td>SSNRSN</td>
<td>$P_\Delta$</td>
<td>AVG</td>
<td>82.7 ± 1.27</td>
<td>77.0 ± 1.22</td>
<td>71.8 ± 1.63</td>
<td>70.7 ± 0.32</td>
<td></td>
</tr>
<tr>
<td>SSNRSND</td>
<td>ES</td>
<td>AVG</td>
<td>42.6 ± 0.72</td>
<td>27.3 ± 1.07</td>
<td>38.2 ± 1.92</td>
<td>22.8 ± 0.33</td>
<td></td>
</tr>
<tr>
<td>SSNRSND</td>
<td>MPS</td>
<td>AVG</td>
<td>45.4 ± 1.15</td>
<td>29.4 ± 0.73</td>
<td>40.8 ± 2.00</td>
<td>24.2 ± 0.26</td>
<td></td>
</tr>
<tr>
<td>SSNRSND</td>
<td>$P_\Delta$</td>
<td>AVG</td>
<td>36.7 ± 1.29</td>
<td>18.6 ± 0.98</td>
<td>34.9 ± 1.82</td>
<td>12.2 ± 0.15</td>
<td></td>
</tr>
<tr>
<td>SSNRSND</td>
<td>ES</td>
<td>NN</td>
<td>68.1 ± 1.52</td>
<td>61.3 ± 1.91</td>
<td>56.3 ± 3.01</td>
<td>54.8 ± 0.45</td>
<td></td>
</tr>
<tr>
<td>SSNRSND</td>
<td>MPS</td>
<td>NN</td>
<td>71.3 ± 1.57</td>
<td>64.8 ± 1.48</td>
<td>59.4 ± 1.65</td>
<td>57.2 ± 0.34</td>
<td></td>
</tr>
<tr>
<td>SSNRSND</td>
<td>$P_\Delta$</td>
<td>NN</td>
<td>70.8 ± 1.53</td>
<td>64.4 ± 1.21</td>
<td>59.1 ± 2.19</td>
<td>57.4 ± 0.42</td>
<td></td>
</tr>
</tbody>
</table>