

Fractional Distance Measures for Content-Based Image Retrieval

Peter Howarth and Stefan Ruger

Department of Computing, Imperial College London,
South Kensington Campus, London SW7 2AZ, UK
{peter.howarth, s.rueger}@imperial.ac.uk

Abstract. We have applied the concept of fractional distance measures, proposed by Aggarwal et al. [1], to content-based image retrieval. Our experiments show that retrieval performances of these measures consistently outperform the more usual Manhattan and Euclidean distance metrics when used with a wide range of high-dimensional visual features. We used the parameters learnt from a Corel dataset on a variety of different collections, including the TRECVID 2003 and ImageCLEF 2004 datasets. We found that the specific optimum parameters varied but the general performance increase was consistent across all 3 collections. To squeeze the last bit of performance out of a system it would be necessary to train a distance measure for a specific collection. However, a fractional distance measure with parameter $p = 0.5$ will consistently outperform both L_1 and L_2 norms.

1 Introduction

The goal of Content-Based Image Retrieval (CBIR) is to provide the user with a way to browse or retrieve images from large image collections, based on visual similarity. At the heart of any CBIR system are visual features that have been extracted from images and distance measures that are used to quantify the similarity between these features. The combination of these two attributes will drive the overall performance of a system.

Visual features are a compact representation of a specific visual facet of an image, such as colour, texture or shape. They are often high-dimensional. Dimensionality of the order of 10^2 to 10^3 is common. Each feature has its own characteristics, such as sparsity, dimensionality and correlation between elements.

A distance (or similarity) measure is a way of ordering the features from a specific query point. These can take many forms. They can be described as a function that maps the \mathbb{R}^n feature space to a one dimensional distance or similarity. The retrieval performance of a feature can be significantly affected by the distance measure used. Ideally we want to use a distance measure and feature combination that gives best retrieval performance for the collection being queried. Often the commonly used distance measures, such as the L-norms, are used as a matter of course. However, a lot can be gained by careful selection of a suitable measure.

In this paper we have applied a fractional distance measure proposed by Aggarwal et al. [1] to the CBIR domain. These measures are an extension of the commonly used L-norm metrics which include Manhattan and Euclidean distance measures. The authors demonstrated that the measures were effective when applied to high-dimensional database vectors for data mining problems, outperforming the more frequently used l_p norms.

This paper is organised as follows. Section 2 discusses the details of fractional distance measures. Section 3 describes how we devised experiments to evaluate the effectiveness of distance measures and Section 4 sets out the results and analysis.

2 Fractional Distance Measures

There are a large number of distance measures that have been used for CBIR. Common ones include: Manhattan, Euclidean, Mahalanobis, and histogram intersection. It is accepted that the choice of proximity measure can have a profound effect on local topology. This is significant for CBIR as when querying a multimedia database we are normally interested in the nearest neighbours. However, often the choice of distance measure is made without much thought. The Euclidean distance metric has its basis in 2 and 3 dimensional space and in this context it is the physical distance measured in a straight line. For higher dimensions it loses its significance, although it is often used as a matter of course.

Beyer et al. [2] set out the problem with nearest neighbour search in high dimensions. That is, that as the dimensionality increases, the distance to the nearest and farthest neighbours tend to converge to the same value. This occurs with most reasonable data distributions and distance measures. The implication of this is that the contrast between data points becomes insignificant as dimensionality increases. Correspondingly, nearest neighbour search may no longer be meaningful. It would therefore appear beneficial if we can use a distance measure that preserves the contrast between data points at higher dimensionality.

The L_p norm is usually induced by the distance,

$$\text{dist}_d^p(x, y) = \left[\sum_{i=1}^d \|x^i - y^i\|^p \right]^{1/p}, \quad (1)$$

where d is the dimensionality of the space and p is a free parameter, $p \geq 1$. Aggarwal et al. [1] extended this definition to allow $p \in (0, 1)$. Please note that strictly speaking the fractional measures defined by dist^p with $p \in (0, 1)$ are no longer distances in the mathematical sense as the triangle inequality is violated. The reason for this is that the a ball with radius one under dist^p is no longer convex for $p < 1$, see Figure 1. This can have an effect on some indexing and partitioning schemes that rely on the metric properties. Nevertheless dist^p still conveys a sense of closeness and we will refer to it as a fractional distance.

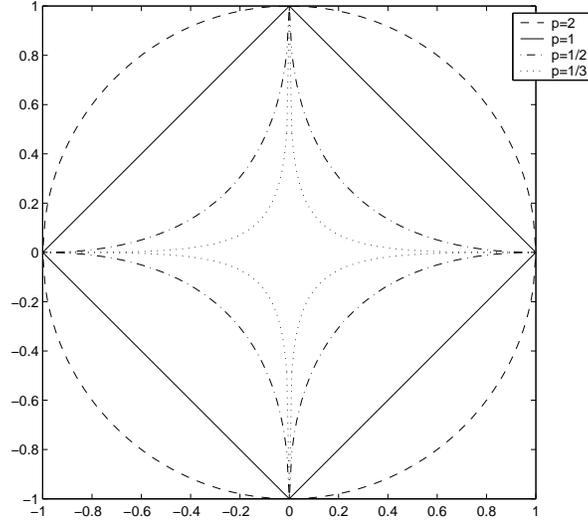


Fig. 1. Unit balls for $p = 2$, $p = 1$, $p = 1/2$ and $p = 1/3$

In [1] a relative distance measure was used to describe the characteristics of the distance space. This had been adapted from [2]; it is defined as:

$$\frac{D \max_d^p - D \min_d^p}{D \min_d^p}, \quad (2)$$

where $D \max_d^p$ is the maximum dist^p between 2 points in a d dimensional distribution, and $D \min_d^p$ is the corresponding minimum distance. This can be used as a measure of the meaningfulness of a distance distribution. In particular [1] showed two results applicable to both ordinary ($p \geq 1$) and fractional ($p \in (0, 1)$) distances.

The first was that the absolute difference between the maximum and minimum distances increases at the rate of $d^{1/p-1/2}$. Thus the smaller the value of p the greater the divergence. Secondly that the relative contrast has the following bounds:

$$C \sqrt{\frac{1}{2p+1}} \leq \lim_{d \rightarrow \infty} E \left[\frac{D \max_d^p - D \min_d^p}{D \min_d^p} \right] \leq C(n-1) \sqrt{\frac{1}{2p+1}}. \quad (3)$$

This is for a uniform distribution of n points and a constant C . It is an interesting result as it shows that fractional measures should have better relative contrast than ordinary distances.

These findings still leave some questions to be investigated. For fractional measures they were based on uniform distributions. They would indicate that the smaller the value of p the better the relative contrast. Whilst this may be the case, with CBIR systems we are interested in the retrieval performance. Altering

the value of p may increase the contrast but could also adversely affect the local neighbourhood and therefore the retrieval performance. In addition the bounds are wide so the nature of the distribution of points may have a significant effect.

One qualitative explanation for the the better performance of L_1 over L_2 is that it is less affected by outliers and therefore noise in high dimensional data. In Euclidean space the distant components will dominate the distance measure. Using L_1 gives near and far components the same weighting. By moving to fractional measures we are adding importance to the components that are similar and removing emphasis from those that are different. This intuitively makes sense as the human visual system can detect small differences in neighbouring patches equally as well as large differences.

3 Experiments

3.1 Overview

The aim of our experiments was to ascertain if fractional distances can be applied to visual features and give an improvement in retrieval performance. Our experiments were designed to address the following questions:

- Do fractional distances increase retrieval performance for high dimensional visual features?
- How does the performance vary with the fractional parameter p ?
- If there is an optimal p for a specific feature is this stable across different image data sets?
- Is it possible to predict the optimal setting for p from any characteristics of the feature or the resultant distance distribution?

We use mean average precision (m.a.p.) as a measure of performance of distance measures. This is because we are interested in performance in the context of a CBIR task. Whilst m.a.p. can be criticised for not being related to a specific user task it does give a good overall measure of performance that trades off between precision and recall. M.a.p. is widely adopted for information retrieval and we therefore feel justified in its use.

3.2 Data Sets

It is recognised with image retrieval that the data set used can have a large influence on results of any experiments and the resultant conclusions. To ensure that our results were not just a feature of the data set used we ran experiments using three different collections. Our primary experiments were done with a collection taken from the Corel image library. These were then followed up with further experiments using TRECVID 2003 and ImageCLEF 2004 collections. This enabled us to validate our results and draw conclusions about the general applicability across three very different collections. The collections and queries are described below.

Corel. We used a subset of Corel that was created by Pickering et al. [3] to evaluate visual features. 6,192 Corel images were carefully selected to give 63 categories that were visually similar internally, but different from each other. This was then split into two sets. The first, a set of 1,548 images, was used to query the remaining 4,644 images. From the query collection we generated single and multiple image queries across all categories. The number of images per query was varied from 1 to 6; for each number we created 630 queries. This made 3,780 in total. The results shown in Section 4 are the mean average precision across these queries.

TRECVID 2003. This collection is widely used. It is much larger than Corel but has drawbacks mainly due to the limited number of queries. It comprises of 32,318 key-frames from TRECVID 2003 video collection [4]. These were taken from ABC and CNN news broadcasts. The search task specified for TRECVID consists of 25 topics. For each topic a few example images were given as a query. The published relevance judgements for these topics were used to evaluate the retrieval performance for different combinations of features and distance measures.

ImageCLEF 2004. This is a medical image collection comprising of 8,725 images, 24 single image queries plus ground truth. It was created for evaluation on the image track of the Cross Language Evaluation Forum [5]. The dataset is quite different to others in that the images are mainly X-rays, CT-scans and medical photographs. The majority of images are monochrome and are carefully posed. It therefore provides an interesting contrast to the other collections.

3.3 Methods

For multiple image queries we used the k -nearest neighbour (k -nn) retrieval approach. Previous work in our group [3] has demonstrated that this outperforms the vector space models for multi-image queries. It is based on the idea that given positive and negative example images, the test images can be classified according to their proximity to these examples. A version of the distance weighted k -nn approach was used [6]. Positive examples (P) are supplied as the query and negative examples (N) randomly selected from the collection. To rank an image i in the collection we identify those images in P and N that are amongst the k -nearest neighbours of i . Using these neighbours we determine the dissimilarity:

$$D(i) = \frac{\sum_{n \in N} (\text{dist}(i, n))^{-1}}{\sum_{p \in P} (\text{dist}(i, p))^{-1}} . \quad (4)$$

A value of $k = 40$ was used for our experiments. A small positive constant value is added to the denominators to prevent division by zero.

3.4 Visual Features

We used a range of high dimensional visual features. These were based on colour, texture and structure. Full details are available in [3, 7]. A brief summary is below:

- **RGB**, this is a joint colour histogram defined in RGB colour-space. It has $8 \times 8 \times 8 = 512$ bins and is sparse.
- **HSV**, this is a joint colour histogram defined in the hue, saturation and value colour-space. The arrangement of bins used is $8 \times 5 \times 5$, giving a relatively sparse 200 dimensional vector.
- **HDS**, this is the MPEG-7 colour structure descriptor. It has 184 non uniformly quantised bins and is relatively sparse.
- **Gabor**, this is a texture feature generated using Gabor wavelets. A bank of 2 by 4 filters are used to detect different scales and directions that characterise a texture. These are applied to image tiles to give additional discrimination. The resultant vector has dimensionality of 560 and is relatively densely populated.
- **Convolution**, this feature discriminates between low level structures in an image. It is created by filtering the image with 25 low level filters designed to detect primitive structures. The resulting feature maps are then re-filtered giving a 625 dimensional feature that is relatively sparse.
- **Thumbnail**, this is created from the pixel intensity values of a scaled down image. We used a size of 40 by 30 resulting in a dense vector of length 1200. This feature is a good discriminator for near identical images.

4 Results

4.1 Performance of Fractional Distances

Corel. The first set of experiments, with the Corel collection, were aimed at determining if fractional distance measures gave a significant retrieval performance gain across a range of visual features. We generated the visual features described in Section 3.4 and ran our query set against these. The results are plotted in Figure 2, which shows mean average precision retrieval against p .

The first thing to note from this graph is that all the features show an increase in m.a.p. for fractional distances. The most significant increases are for the RGB, HSV, HDS and convolution features. The Gabor and thumbnail features are both flat across the graph, showing only a slight improvement in retrieval performance for fractional distances. The position of the maxima vary from feature to feature but all fall between p values of 0.25 and 0.75.

The HDS feature shows the maximum relative gain in m.a.p.. It increases from 18.2% at $p = 1$ to 23.6% at $p = 1/4$, a relative gain of 30%.

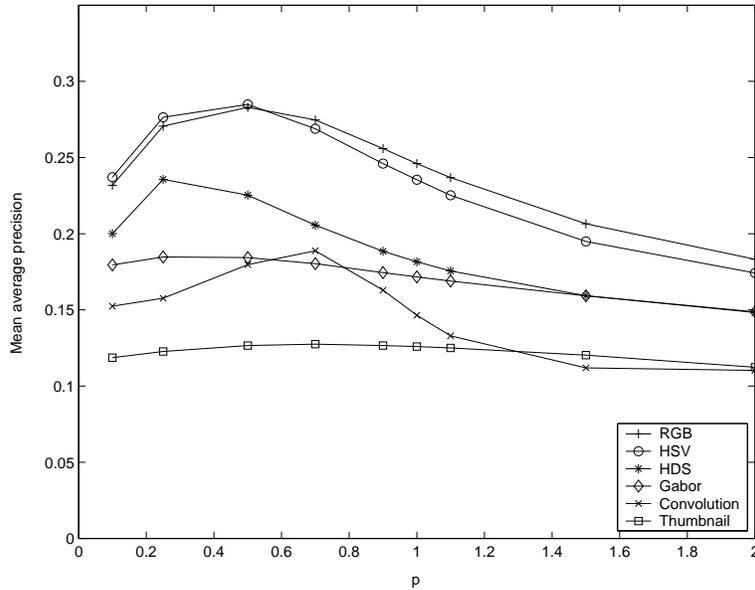


Fig. 2. Graph of retrieval results for Corel

TRECVID 2003. The larger TRECVID collection presents more of a challenge for image retrieval. We generated the same features as for Corel. The retrieval performance is shown in Figure 3. The results show a marked performance increase for fractional distance measures.

The overall results are very similar to those for Corel. RGB, HDS, HSV and convolution features show increased performance for fractional distances. Similarly, the performance for the Gabor and thumbnail features does not improve. The maximum gain in m.a.p. is shown by the RGB feature which increases from 2.0% at $p = 1$, to 3.3% at $p = 1/2$. This is a relative increase of 65%.

The plots from the 2 experiments have the same characteristic shape, with the maxima falling between 0.25 and 0.75. However, a detailed examination of the p values at maximum retrieval for each feature shows that they are different to the Corel collection. This demonstrates that the optimum value of p is not independent of the data collection.

ImageCLEF Fewer colour features were used with the ImageCLEF collection due to its mainly monochrome nature. The results are plotted in Figure 4.

Examining the ImageCLEF results we can see that the general trend is similar to those from Corel and TRECVID. HDS and convolution features show performance gains for fractional p values. The convolution feature has a much larger relative gain than for TRECVID whereas HDS only has a slight gain. The performance of the Gabor feature reduces for fractional p values. The significant

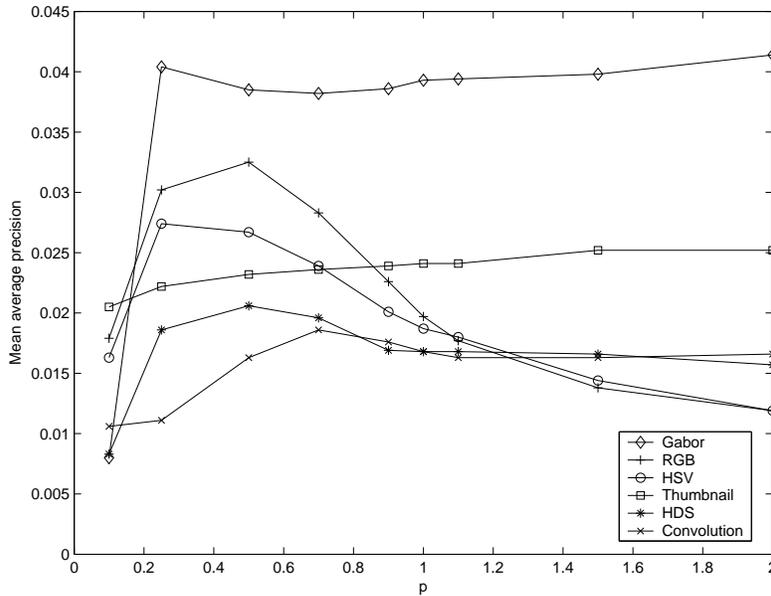


Fig. 3. Graph of retrieval results for TRECVID2003

difference in the results is for the thumbnail feature. In contrast to the 2 previous experiments it shows a marked performance gain for fractional distances.

To explain these results we must consider the characteristics of the collection. It contains a large proportion of monochrome images and because of the medical subject contains groups of near identical images. For example X-rays of a specific part of the body will always be composed in exactly the same way. In addition the queries for this collection are single images.

The effect of the monochrome images on colour features will be to reduce the dimensionality. Qualitatively this explains the reduced gain for the HDS feature. The increase in performance for thumbnail may be due to the groups of very similar images in the collection. This feature will discriminate these effectively and it appears that its performance is enhanced by the fractional distance measure.

4.2 Discussion of results

Overall the results on Corel, CLEF and TRECVID show that the performance benefits of fractional distance measures are generally applicable across widely differing datasets, features and queries.

All the features, except Gabor and thumbnail, consistently show an increase in retrieval performance when used with fractional distance measures. The max-

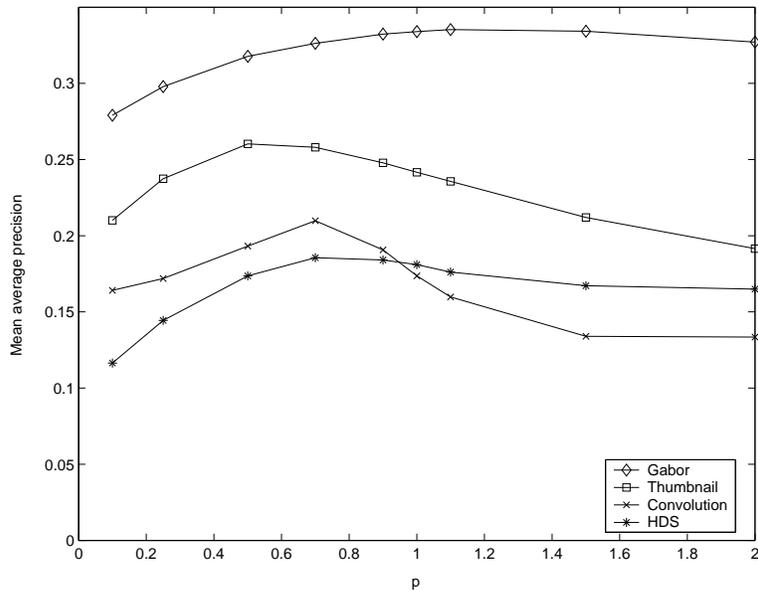


Fig. 4. Graph of retrieval results for ImageCLEF2004

imum gains appear at values of p between 0.25 and 0.75. The optimum value of p varies depending on the combination of feature and test collection.

In an attempt to find a predictor for the optimum value of p we investigated the statistical properties and dimensionality of the space defined by the features and test collection. No clear relationship was found. We intend to research this further.

Taking a more qualitative viewpoint, the 2 features that do not respond well to fractional distances are both dense vectors. The features with the greatest improvement are all sparse vectors. It would therefore appear that the sparsity of the feature vector may be a general indicator that use of a fractional distance measure will improve mean average precision retrieval.

Intuitively this makes sense as fractional distance measures give more weight to element comparisons where the values are similar, i.e. 2 zeros, or 2 non-zero values. With sparse features a large number of element-wise comparisons will be between zero and some value. The contribution of these to the total distance will add noise that may swamp the overall similarity. These will be given less importance with fractional distances than with higher norms.

5 Conclusion

We have shown that fractional distance measures give a significant improvement in mean average precision retrieval over the commonly used L_1 and L_2 norms.

The performance gains were consistent when using high dimensional visual features over three different image collections.

By experimenting across very different data sets we have shown that the optimum value of value of p for a feature cannot be determined by training on a single collection. It is linked to the combination of both feature and dataset. However, we have demonstrated that a choice of $p \in (0.25, 0.75)$ improves mean average precision across nearly all features and datasets. To find the optimum p the distance measure would need to be learnt for each collection. However, a value of $p = 0.5$ will improve retrieval performance in nearly all circumstances.

We could not determine a reliable predictor for the optimum value of p . However, qualitatively there appears to be a link between the sparsity of the feature vector and how much a fractional distance measure improves retrieval performance. We intend to investigate this further.

References

1. Aggarwal, C.C., Hinneburg, A., Keim, D.A.: On the surprising behavior of distance metrics in high dimensional space. *Lecture Notes in Computer Science* **1973** (2001) 420–434
2. Beyer, K., Goldstein, J., Ramakrishnan, R., Shaft, U.: When is “nearest neighbor” meaningful? *Lecture Notes in Computer Science* **1540** (1999) 217–235
3. Pickering, M., Rüger, S.: Evaluation of key-frame based retrieval techniques for video. *Computer Vision and Image Understanding* **92** (2003) 217–235
4. A Smeaton, W.K., Over, P.: TRECVID 2003 — An introduction. In: *TRECVID 2003 Workshop*. (2003) 1–10
5. Clough, P., Müller, H., Sanderson, M.: The CLEF Cross Language Image Retrieval Track (ImageCLEF) 2004. In Peters, C., Clough, P., Gonzalo, J., Jones, G., Kluck, M., Magnini, B., eds.: *Fifth Workshop of the Cross-Language Evaluation Forum (CLEF 2004)*, *Lecture Notes in Computer Science (LNCS)*, Springer, Heidelberg, Germany (in print) (2005)
6. Mitchell, T.M.: *Machine Learning*. McGraw Hill (1997)
7. Howarth, P., Rüger, S.: Evaluation of texture features for content-based image retrieval. In: *Proceedings of the International Conference on Image and Video Retrieval*, Springer-Verlag (2004) 326–324

Acknowledgements: This work was partially supported by the EPSRC, UK.