Very Fast Image Retrieval

Diogo André da Silva Romão

Abstract
Nowadays, multimedia databases are used on several areas. They can be used at home, on entertainment systems or even in professional context as medical information system. In order to retrieve stored information from these systems, specially images, the user must be aware on what will the retrieval be based on, either on text or on content. Content based image retrieval (CBIR) is a technique which allows the user to retrieve images based on its content rather than on its metadata, such as keywords, or descriptions associated with the image. The term “content” in this context may refer to colour, shape, texture or any other information that can be derived from the image itself. The main goal is, therefore, to retrieve the most similar image (images) given a certain object, called query object. In order to do so, it is necessary to use several resolutions for the images present on the database where a multimodal algorithm is called and where images are grouped into small groups. For each group, the most similar images to the given query object are computed. This algorithm is executed repeatedly, on a higher dimensional space at each iteration, until the most similar image to the given query object is found.

Keywords: Curse of Dimensionality, Content Based Image Retrieval (CBIR), High-Dimensional Indexing, Hierarchical Linear Subspace Method (HLSM), multimodality

1. Introduction
Over the last years, with smartphones, personal computers and tablets mass usage, an increase in multimedia documents has been observed. With technological advancements, the amount of information present on these objects grows bigger. 1080p and 4K content are examples of these technological breakthroughs, which show no sign of stopping, having been recently announced 8K content. While the information on these objects increased, so did their size, meaning that, existing algorithms to retrieve these objects became slower.
Multimedia documents can be used at home, for instance on photos and videos but they can also be used in medical examination (Loening et al., 2003), or even by NASA on Mars (Rui et al., 1997). Bearing this in mind, it is extremely important that retrieval solutions have the ability to perform a fast retrieval event when these objects size increases.
The vast majority of existing solutions show an exponential decrease in performance when dimension arises, meaning those solutions may no longer be used. There are a few solutions which are able, up to a certain point, to bear with increased dimension but have other limitations. These limitations have to do with the fact that these solutions are only able to deal with one modality. Therefore, there is a need for a new solution, capable of dealing with high-dimensional objects as well as multimodality.

2. Literature review
The proposed methodology combines multimodal search with the Hierarchical Linear Subspace Method (Wichert., 2008).

3. Multimodality
Given that, the final algorithm must be based on multimodality, then multimodality must be part of the solution. Nevertheless, it is mandatory to understand which type of multimodal search (early fusion or late fusion) gets the best results. Late fusion is often chosen as the best choice for multimodal search (Snoek et al., 2005), but that doesn’t mean that, in certain cases, we won’t
obtain better results through early fusion. This means that, in order to understand which kind of multimodal search best applies to this case, the algorithm must be tested with both early and late fusion.

3.1. Hierarchical Linear Subspace Method

The Hierarchical Linear Subspace Method is a method which is able to avoid curse of dimensionality’s (Volnyansky et al., 2009) consequences. This is accomplished through the usage of subspaces. Each subspace represents an object resolution. To be possible to use HLSM all multimedia objects present must be resized to the resolutions corresponding to each subspace.

This is mandatory because, when trying to retrieve the most similar image present in the database with a given input image, we first start by comparing these images on their lower available resolution, meaning that, we will primarily compare them on the lower subspace. At each subspace the input image must be compared to all the existing images in the database. If their distance is bigger than a given value ε, threshold, than the image is discarded. If the distance between an input image and an image present on the database is smaller than the threshold than the images must be compared on the next subspace (Wichert., 2008). The process is repeated until there are no more images to be considered or all similar images have been compared in the original subspace level.

Looking at figure 1, it is possible to observe the database’s characteristic using this method.

This happens when the fifty images retrieved from the totality of the database are the same as when only two thousand and five hundred images are used. Nevertheless, this worst case scenario is extremely rare, meaning that, when increasing the number of used images, new similar images to the input image will be included. This way, the threshold will decrease. Decreasing the threshold results in a decreased number of operations needed resulting in a faster search. This way, it is clear to see that if a method works for a higher threshold it will also work for a lower threshold and produce better results.

4. The methodology

In order to validate the solution, several methods, all using HLSM, were developed and tests were made for each one.

For the first eleven methods, two unimodal methods and nine multimodal methods, 2500 images were used, about a fourth of the complete database. There was no need to use the totality of the database due to the fact that, when increasing the number of imagens, the threshold decreases, meaning that, in the worst case scenario, the threshold maintains its value.
Observing the figure, it is possible to state that, in order to retrieve five hundred images, a threshold of $1.5 \times 10^4$ is needed.

At the table below, it is possible to see the number of operations and the threshold needed to retrieve ten, twenty, thirty, forty and fifty images respectively.

Table 1 – Results obtained from the tests to the unimodal method a)

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED</th>
<th>THRESHOLD</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>8604.2</td>
<td>24753448</td>
</tr>
<tr>
<td>20</td>
<td>9005.7</td>
<td>29008516.3</td>
</tr>
<tr>
<td>30</td>
<td>9271.6</td>
<td>32122399.7</td>
</tr>
<tr>
<td>40</td>
<td>9469.3</td>
<td>34519131.4</td>
</tr>
<tr>
<td>50</td>
<td>9645.4</td>
<td>36695240.6</td>
</tr>
</tbody>
</table>

Looking at the table above, it is possible to state that the bigger the number of retrieved images, the bigger the number of needed operations. This happens due to the fact that, in order to retrieve more images, a bigger threshold is needed. Using a bigger threshold, images that were dissimilar may now be similar, resulting in additional comparisons on the higher dimension subspaces, resulting in an increased number of operations.

4.1.2. Method b)

The second unimodal method used, method b) is pretty similar to the first one, being the only difference the modality used. In this method, the used method is the image histogram.

The following image shows the database’s characteristic using the method b).

![Figure 2 – Unimodal method b) database’s characteristic for 2500 images.](image)

Analysing the image above it is possible to observe that, in order to retrieve five hundred images, the needed threshold is much smaller than it was using the unimodal method a).

The following table shows the number of operations, the threshold and the respective number of retrieved images when using the unimodal method b).

Table 2 – Results obtained from the tests to the unimodal method b)

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED</th>
<th>THRESHOLD</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1253</td>
<td>283205.8</td>
</tr>
<tr>
<td>20</td>
<td>1369.1</td>
<td>317197.4</td>
</tr>
<tr>
<td>30</td>
<td>1449.3</td>
<td>353777.3</td>
</tr>
<tr>
<td>40</td>
<td>1510.3</td>
<td>392255.4</td>
</tr>
<tr>
<td>50</td>
<td>1553.7</td>
<td>424216.3</td>
</tr>
</tbody>
</table>

Through the table reading, and as expectable, it is possible to state that the average number of operations needed to retrieve the same number of images as in the unimodal method a) is now much smaller. This is due to the fact that the feature vectors resulting from the image histogram are much smaller than the feature vectors resulting from the image as a matrix. This is given to the fact that the feature vector for the image histogram only has information about the image colour while the feature vector for the image as a matrix has information about the image’s shape as well as the colour of each pixel of the image. This results in a smaller number of needed operations when comparing images.

4.2. Multimodal methods using early fusion

As previously explained, the image as a matrix modality is the one with the most information about the image. This way, when combining both modalities, it will be possible to obtain data about images’ shape as well as images’ colour. Given this, all multimodal methods will use the unimodal method a) as reference since it has the same data about these images.

4.2.1. Method c)

The method c) is the first to use multimodality. Being the first, it was intended to be as simple as possible to be used as a starting point to all the following multimodal methods using early fusion.

This way, given that there are only four histogram subspaces, only four subspaces for the image as a matrix were used.

Once this method uses early fusion, data fusion is needed beforehand and each modality must have its own weight. For this method, a weight of 50% was chosen for each of the modalities.

Given that, four subspaces are used for each modality, in order to fuse modalities, the same subspaces must be joint. This means that, the lower subspace feature vector for the image as a matrix modality must be combined with the lower
subspace feature vector for the histogram modality.

Computing the distance between images through multimodality using early fusion is similar to when unimodality is used. In this case, two distances between the images are computed, for each modality. After that, each distance is multiplied by the weight of the respective modality. Adding both results, we get the distance between the images. When using HLSM, this process must be applied on every subspace. Images considered dissimilar (distance bigger than the threshold) are discarded and similar images must be compared on the next subspace.

Figure 3 shows the result of the characteristic's computation for the database given the present method.

![Figure 3](image)

Figure 3 – Multimodal method c) database's characteristic for 2500 images.

Through the observation of the previous figure, it is possible to state that the threshold needed to retrieve the same number of images when comparing to the unimodal method a) is smaller. This happens because the threshold value for the image histogram is really low when comparing with the same value for the image as a matrix modality. Given this, when combining modalities, a lower threshold is obtained than when no combination is done. In this specific case, when comparing unimodal methods with multimodal methods, a lower threshold doesn’t imply a lower number of operations. This is clearly to see when looking at the table below and comparing its values to table 1. In this case, the threshold to retrieve the same amount of images is lower but the number of required operations is bigger.

### Table 3 – Results obtained from the tests to the multimodal method c)

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED</th>
<th>THRESHOLD</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5425,3</td>
<td>2739967,4</td>
</tr>
<tr>
<td>20</td>
<td>5674,8</td>
<td>31878952,3</td>
</tr>
<tr>
<td>30</td>
<td>5826,9</td>
<td>34854069,1</td>
</tr>
<tr>
<td>40</td>
<td>5933,2</td>
<td>37115093,8</td>
</tr>
<tr>
<td>50</td>
<td>6025,6</td>
<td>39106681</td>
</tr>
</tbody>
</table>

A possible cause for this worsening may be the fact that there were available subspaces for the image as a matrix modality that were not used. Through the usage of all subspaces it is expected the number of operations to drop. This hypothesis is tested on the following method.

**4.2.2. Method d)**

For the fourth method, the second multimodal, the totality of the subspaces for both modalities were used in an attempt to reduce the average number of operations.

Having more subspaces for the image as matrix modality than for the image as histogram, there’s the need for a new way to combine subspaces. This way, the sixth subspace for the image as a matrix will be combined with the fourth subspace for the image as histogram. Afterwards, the fifth subspace for the image as a matrix will be combined with the fourth subspace for the image histogram. After that, the combination will follow the logic used on method c).

Like in the previous method, a weight of 50% was assigned for each of the modalities.

Looking at figure 4, it is possible to state that the threshold needed to retrieve the same number of images as on method c) is equal.

![Figure 4](image)

Figure 4 – Multimodal method d) database's characteristic for 2500 images.

This happens because the distance between images is given by their distance on the original subspace. Since these two methods only differ
on the lower subspaces, no changes on the threshold value are observable. Although there’s no change on the threshold value, a change for the number of operations is expected. This happens due to the addition of new subspaces. This way, it is possible to discard images earlier and consuming less operations.

Looking at the following table, it is possible to verify that, comparing both methods c) and d), although both have the exact same threshold values, the number of operations required to retrieve the same number of images is lower on method d). Despite having reduced the number of operations, this remains higher when comparing to the unimodal method a).

Table 4 – Results obtained from the tests to the multimodal method d)

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED</th>
<th>THRESHOLD</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5435,3</td>
<td>27070047,8</td>
</tr>
<tr>
<td>20</td>
<td>5674,8</td>
<td>31640563,7</td>
</tr>
<tr>
<td>30</td>
<td>5626,9</td>
<td>34672853,8</td>
</tr>
<tr>
<td>40</td>
<td>5933,2</td>
<td>36973453</td>
</tr>
<tr>
<td>50</td>
<td>6025,6</td>
<td>38992241,8</td>
</tr>
</tbody>
</table>

4.2.3. Method e)

Given that using the previous method was not possible to improve the number of operations when comparing with the method a), a new approach was created.

This method, like method d), will use all subspaces from all modalities. The big difference will be the weights assigned to each modality. This way, it is expected a drop for the number of operations when comparing to the previous approach.

Given that the image as a matrix modality has more information about the image that the image histogram, a bigger weight will be assigned to the former. This way, a weight of 80% will be given to the image as a matrix and 20% will be assigned to the image histogram.

Through the following figure, it is possible to observe the database’s characteristics for the present method.

Comparing both the previous and the present method, it is possible to see that the present one needs a higher threshold in order to retrieve the same number of images. This is due to the fact that the number of operations is higher for the image as a matrix modality than it is for the image histogram. This way, given that the image as a matrix modality was given a higher weight than it was on the previous method, the threshold is higher.

Observing the following table, it is possible to verify that although the threshold is bigger using method e), when 20 images are retrieved, the number of operations is lower when comparing to the method d). This improvement is due to the fact that the images retrieved in both method d) and e) are not the same. This way raising the weight for the image as a matrix modality results in images that were dissimilar being even more dissimilar. This way it is possible to eliminate these images in an earlier stage.

Despite this, the number of operations is still higher than it was for the unimodal method a).

Table 5 – Results obtained from the tests to the multimodal method e)

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED</th>
<th>THRESHOLD</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7421,5</td>
<td>27259631</td>
</tr>
<tr>
<td>20</td>
<td>7738,4</td>
<td>31601300,6</td>
</tr>
<tr>
<td>30</td>
<td>7952,7</td>
<td>34674492,5</td>
</tr>
<tr>
<td>40</td>
<td>8111,2</td>
<td>37015586,9</td>
</tr>
<tr>
<td>50</td>
<td>8246,9</td>
<td>39146389,9</td>
</tr>
</tbody>
</table>

4.2.4. Method f)

Given the fact that it hasn’t yet been possible to obtain improvements through multimodality when comparing to the unimodal method a), new possibilities must be taken into account. Observing the database’s characteristic for the image histogram, it is possible to see that the
loss of information between subspaces is low. This is the starting point for the next method. Taking this into account only the lower subspace for the image histogram is used for the present method while the image as a matrix modality will have all its subspaces used. This means that, each subspace for the image as a matrix modality will be combined with the lowest subspace for the image histogram.

Comparing the last two methods, method d) and method e), it is possible to state that the best results in terms of number of operations were obtained when both modalities had the same weight. This way, for the present method, both modalities have been assigned with a weight of 50% each.

Through this method it is expected to obtain significant improvements comparing to the multimodal methods previously used. Looking at the next figure it is possible to see the database’s characteristics for the present method.

![Multimodal method f) database’s characteristic for 2500 images.](image)

As it is possible to state, comparing the method’s results with method’s e), a significant drop for the number of operations is observed. This is clearer when the number of images retrieved is higher. This happens because, on higher subspaces, the lowest subspace for the image histogram is used, meaning that a lower number of operations will be performed than it was before. This way, the higher the subspace and the more images retrieved, the better the results obtained when comparing with the previous multimodal methods.

Comparing the results obtained from this method with the results obtained from the unimodal method a), it is possible to see that no improvement whatsoever has been achieved. Despite this, the number of operations is now pretty similar between both methods.

**4.2.5. Method g)**

As it has been possible to verify, none of the previous multimodal methods allowed an improvement in terms of number of operations when comparing with the unimodal search. This happens because early fusion is used on all multimodal methods tests. Through early fusion, the combination of feature vectors takes place before the search, meaning that, the resulting feature vector, on which the search will be performed, will have a bigger dimension. This way a bigger number of operations will be needed when retrieving images. Given this, in order for multimodal methods using early fusion to work, the resulting feature vector from combining the modalities needs to be smaller.

Attending to the modalities in question it’s possible to state that both contain the same information. Through the image as matrix modality it is possible to know the colour of each image’s pixel. Through the image histogram is possible to know the probability of a given colour being present in the image. Therefore, this information is redundant.

In order to aim for a smaller resulting feature vector, for the present method all images were converted to grey scale. This way, combining the image as a matrix for the grey scale image and the image histogram for the colourful image, it is expected that their combination will result on a smaller feature vector, reducing even further the number of operations necessary to retrieve images.

For this first method using this approach all subspaces for the image as matrix modality were used. On each level the feature vector corresponding to this modality will be combined...
with the image histogram feature vector corresponding to the original subspace. This way, if an improvement is achieved, certainly it can be better through HLSM on both modalities. A weight of 50% was assigned to both modalities.

Figure 7 shows that the threshold to retrieve the same amount of images when comparing to all the previous multimodal methods is even lower than the threshold needed to retrieve the same amount of images when using the unimodal method a).

![Figure 7](image7.png)

**Figure 7 – Multimodal method g) database’s characteristic for 2500 images.**

The following table shows the number of operations needed to retrieve the same amount of images as in the methods before.

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED</th>
<th>THRESHOLD</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3005,8</td>
<td>2388836,5</td>
</tr>
<tr>
<td>20</td>
<td>3443,3</td>
<td>2587846,6</td>
</tr>
<tr>
<td>30</td>
<td>3532,1</td>
<td>2862026,2</td>
</tr>
<tr>
<td>40</td>
<td>3594,7</td>
<td>3122636,3</td>
</tr>
<tr>
<td>50</td>
<td>3647,5</td>
<td>3373204,5</td>
</tr>
</tbody>
</table>

As it is possible to verify, this method is the one that produces best results amongst the multimodal ones. It produces even better results than the multimodal method a) allowing a reduction of 11 million when 10 images are retrieved and an even better reduction when more images are retrieved.

As stated before, this results can be even better since no HLSM was used for the image histogram modality.

### 4.2.6. Method h)

For the present method, as it happened on the previous one, HLSM will be used for the image as a matrix. Unlike the previous one, HLSM will also be used for the image histogram modality. A weight of 50% was assigned to each modality. Given that HLSM was introduced to the image histogram, it is expected that method h) will generate better results than multimodal method g).

Comparing the database’s characteristics for method g) and method h), it is possible to state that both have exactly the same characteristics.

![Figure 8](image8.png)

**Figure 8 – Multimodal method h) database’s characteristic for 2500 images.**

This happens because the threshold is given by the image distance in the original subspace. Since HLSM is used for both modalities, in the original subspace the higher dimension feature vector of both modalities is used, just like the previous method. This way both methods are the same on the higher subspace but not on the lower subspaces. This way, the threshold value remains unchanged.

The following table shows the results for the present method. As expected, using HLSM for both modalities resulted in a drop for the number of operations.

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED</th>
<th>THRESHOLD</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3005,8</td>
<td>10881876</td>
</tr>
<tr>
<td>20</td>
<td>3443,3</td>
<td>12622621</td>
</tr>
<tr>
<td>30</td>
<td>3532,1</td>
<td>13727465,3</td>
</tr>
<tr>
<td>40</td>
<td>3594,7</td>
<td>14535381,6</td>
</tr>
<tr>
<td>50</td>
<td>3647,5</td>
<td>15218211,8</td>
</tr>
</tbody>
</table>

Comparing this method with method a) it is possible to state that the number of operations needed to retrieve 50 images dropped from roughly 37 to 15 million, which represents a 59% reduction. Despite this, from the 50 images retrieved through method a) only 21 of these are retrieved through method h). This means that more than half of the images that were previously similar are now dissimilar.
4.2.7. Method i)

Taking into account that more than half the images that were similar using the unimodal method a) are dissimilar when using multimodal method h), it is of the utmost importance minimizing this difference.

It is intended to speed up the process of retrieving similar images while trying to retrieve the same images as on the unimodal method a). As the goal of reducing the number of operations, the following methods will use the same approach as method h). The differences between method h) and the following will be modalities’ weight. For the present method a weight of 80% was assigned to the image as a matrix modality and 20% was assigned to the image histogram. This way, it will be possible to understand which of the modalities is of higher importance. If the results are better than the ones obtained through method h) image as a matrix is more important than the image histogram. Otherwise, image histogram is more important.

Through the observation of the database’s characteristics for the present method is possible to verify that the number of operations to retrieve the same number of images increased when comparing with multimodal method h). Despite this increase it is still below from the values observed for the unimodal method a).

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED</th>
<th>THRESHOLD</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>410,3</td>
<td>12409360,5</td>
</tr>
<tr>
<td>20</td>
<td>427,1</td>
<td>13945576,2</td>
</tr>
<tr>
<td>30</td>
<td>4370,1</td>
<td>14911669</td>
</tr>
<tr>
<td>40</td>
<td>4448,5</td>
<td>15675719,5</td>
</tr>
<tr>
<td>50</td>
<td>4511,8</td>
<td>16279261,4</td>
</tr>
</tbody>
</table>

As it is possible to state, as expected the number of operations is higher when comparing with the multimodal method h) but lower in comparison with the unimodal method a). For the present method an increase in the average number of images simultaneously retrieved between methods i) and a) increased from 21 to 31, which turns out to be a better than what was achievable through method h). This method also shows that image as matrix modality for a grey scale image is more important than the image histogram.

4.2.8. Method j) and k)

Since increasing the weight for the image as matrix modality resulted in a higher number of simultaneously retrieved images, the next two methods will increase even further this modality’s weight. The first one, method j), assigns a weight of 95% to this modality, while method k) assigns a 100% weight.

Observing the database’s characteristics for both methods (figure 10 and 11), it is possible to state, as expected, that the threshold will be higher than it was for the multimodal method i).

![Figure 9 – Multimodal method i) database’s characteristic for 2500 images.](image)

![Figure 10 – Multimodal method j) database’s characteristic for 2500 images.](image)
The following tables show the results of the tests made for each of the methods.

**Table 10 – Results obtained from the tests to the multimodal method j)**

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED (IMAGES)</th>
<th>THRESHOLD (OPERATIONS)</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4392.8</td>
<td>12025592.1</td>
</tr>
<tr>
<td>20</td>
<td>4562.7</td>
<td>13455967</td>
</tr>
<tr>
<td>30</td>
<td>4677.2</td>
<td>14115298</td>
</tr>
<tr>
<td>40</td>
<td>4763.7</td>
<td>15147472.3</td>
</tr>
<tr>
<td>50</td>
<td>4842.7</td>
<td>15800236.2</td>
</tr>
</tbody>
</table>

**Table 11 – Results obtained from the tests to the multimodal method k)**

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED (IMAGES)</th>
<th>THRESHOLD (OPERATIONS)</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4461.8</td>
<td>11751313.4</td>
</tr>
<tr>
<td>20</td>
<td>4644.7</td>
<td>13195641.6</td>
</tr>
<tr>
<td>30</td>
<td>4763.4</td>
<td>14162847.8</td>
</tr>
<tr>
<td>40</td>
<td>4855.7</td>
<td>1495871.7</td>
</tr>
<tr>
<td>50</td>
<td>4939</td>
<td>1557062.5</td>
</tr>
</tbody>
</table>

Through these method was possible to retrieve 30 images that were also retrieved using unimodal method a), which is worse than the multimodal method i). The number of needed operations to retrieve this images is also higher than it was on method i). This way, it is possible to state that, method i) is the best method to use when a fast multimodal retrieval is in order.

**4.3. Multimodal methods using late fusion**

The next method’s purpose is to verify whether early or late fusion produce the best results. In order to do so, and according to late fusion, two unimodal searches are performed, one for each modality. The results obtained from these searches are then fused using intersection. In this case, the number of operations is given by the sum of each unimodal method’s number of operations. When intersecting the 50 images retrieved from each unimodal search, the result set only contains, on average, 5 images. This means that, in order to retrieve 5 images using late fusion, two searches need to be performed, each retrieving 50 images. The average number of operations needed to retrieve 5 images using late fusion is 37119465.9, which is higher than the cost of retrieving 50 images using early fusion. This way, is possible to state that, in this case, early fusion performs better than late fusion.

**4.4. Final Test**

Given that multimodal method i) was the one which provided the best results, it’s interesting to see its behaviour when the number of used images grows. Therefore, for this final test the whole amount of images present in the database is used. The method is exactly as described on section 4.2.7.

The following table shows the results obtained from this final test.

**Table 12 – Results obtained from the tests to the multimodal method i) using 9876 images**

<table>
<thead>
<tr>
<th>IMAGES RETRIEVED (IMAGES)</th>
<th>THRESHOLD (OPERATIONS)</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3985.1</td>
<td>38344805.8</td>
</tr>
<tr>
<td>20</td>
<td>4113.6</td>
<td>42556009.4</td>
</tr>
<tr>
<td>30</td>
<td>4192.8</td>
<td>45286891.7</td>
</tr>
<tr>
<td>40</td>
<td>4250.3</td>
<td>47324917</td>
</tr>
<tr>
<td>50</td>
<td>4297.9</td>
<td>49041316.9</td>
</tr>
</tbody>
</table>

As it is possible to see, increasing four times the set of images, results in an increase by three times for the number of operations. This way it is possible to guarantee that the method will work even when the number of images present in the database increases.

**5. Conclusions**

In this paper, several unimodal and multimodal methods for image retrieval are analysed and compared. Through these methods’ results it’s possible to state that, when using image as matrix and image histogram modalities, a multimodal search, using early fusion, can be faster than a unimodal one. It is also possible to state that, for the given database, an early fusion method presents better results than a late fusion one. As a future work, it would be interesting to have methods using the L1 norm instead of the L2 used on all presented methods. It would also be of interest to explore modalities not explored in this work, in order to find out if the proposed...
methods show the same improvement when dealing with other modalities.

6. References


