A FFT based technique for image signature generation

Augusto Celentano[a] and Vincenzo Di Lecce[b]

[a]Università Ca'Foscari di Venezia, Dipartimento di Matematica Applicata e Informatica
Ca'Dolfin, Dorsoduro 3825/e, 30100 Venezia, Italia, auce@dsi.unive.it

[b]Politecnico di Bari, Dipartimento di Elettrotecnica e Elettronica,
Via Orabona 4, 70125 Bari, Italia, dilecce@poliba.it

ABSTRACT

In this paper we address image retrieval by similarity in multimedia databases. We discuss the generation and use of signatures computed from image content. The proposed technique is not based on image annotation, therefore it does not require human assistance.

Signatures abstract the directionality of image objects. They are computed from the image Fourier transform, and the influence of computation parameters on signature effectiveness is discussed.

Retrieval is based on spectrum comparison between a reference image, assumed as the query, and the images in a collection. We introduce a metric for comparing the spectra and ranking the result, and approach the issue of partial query specification. Sample results on a small test collection are given.

Keywords: Fast Fourier Transform, Image feature extraction, Image signature, Content-based retrieval

1. INTRODUCTION

Retrieval by image content has received great attention in recent years. Several approaches have been proposed to the problem of identifying image features that could be used as objective indicators of their contents. Approaches can be divided in two main sets:

- The ones requiring explicit image annotation, made by a human who interprets the content and marks objects, regions, shapes, foreground and background, and so on.
- The ones based on automatic extraction of physical properties like colour, contour and texture, by computing values and distributions that are used in image classification and comparison in exact and approximate retrieval.

Both approaches have strong and weak points: annotation results in accurate interpretation of image meaning, but requires human effort. Automatic features computation can be done in unassisted runs, but hardly relates to real image meaning, except for specific domains and applications. With different targets, a debate on the two approaches may mirror the one on automatic versus manual indexing of text documents.

Automatic feature extraction in many cases may provide good performance. Due to its unassisted nature, it is worth to be considered in large databases, and for applications where the cost of images processing must be kept low in terms of human resources.

In this paper we address a specific area of application, that concerns retrieval by similarity in thematic databases. We define more precisely what we mean with these two terms.

- Retrieval by similarity means that we consider a query that is made of an image, or of one or more parts of an image. The retrieval process consists in retrieving other images that are similar to the query. In other words, some selected visual properties of the retrieved images have values near to the query's ones, according to a suitable metric.

  It is usual to consider information retrieval a non exact process, returning ranked results, without an *a priori* fixed border between the relevant and not relevant items. We follow this schema.
- A thematic database is a database whose content is restricted to a domain. The statement is in some way ambiguous, because it does not define the scope of the domain. A thematic database may thus collect images of very different contents.

For the purpose of our investigation, we assume that the thematic nature allows queries to be targeted to visual aspects rather than to content interpretation. In other words, we assume that the purpose of querying the database, i.e., to select images, is not to identify specific contents, but rather to select close visual appearance.

In order to accomplish the task of retrieving images by visual similarity, we have to define three items:

- An image surrogate, that holds values computed from image features in such a way that simple matching procedures can be used for retrieving images with similar features. Such a surrogate is named, in traditional information retrieval jargon, a *signature*[*].
- A metric, that allows comparison of different signatures (the one of the query against the ones of the database images), possibly according to some filtering procedure that accounts for partial query specification or weight of query terms
- A verification procedure that validates the retrieval made on signature against true image features. It is well known that, being a signature a compact surrogate of the original document (the image), false drops can be generated. We shall face this issue in Section 6.

The paper is organised as follows. In Section 2 a brief overview of the current literature is given. Section 3 discusses the use of lines orientation as a means for matching similar images. Section 4 describes the signature generation technique, and Section 5 presents a retrieval technique based on signature similarity. In Section 6 the proposed approach is discussed. Conclusions are given in Section 7. Acknowledgements and references follow.

2. RELATED WORK

Multimedia information systems (MMIS) and content based information retrieval systems (CBIRS) are manage multimedia information with specific functions for content classification, analysis and retrieval [2,3,4]. A way to automatically compare images is the extraction of features (colour, texture, shape, position) from images, their quantization, and their use as indexes in classification and retrieval.

QBIC [5] is the most representative system that allows content based queries to be performed on shape, texture, colour, and sketch. It is based on image annotation at database population phase for some features like shape and foreground/background identification, thus requiring human assistance. Query results are ranked according to a similarity metric.

Chabot [6] performs content retrieval based only on colour, relying on a textual description for the content identification. It does not rank the results, returning a flat set of images that the user can browse.

ICARS [7] extends text based indexing techniques to the image domain. It retrieves images based on similarity of content using index terms, text description, and user-generated images as a query. The system does not perform object recognition or image segmentation, but relies on a learning process about the position of objects(called atoms) in the image.

Query by texture is discussed in [8,9]. The first paper identifies textures through a number of features like coarseness, contrast and directionality. The second paper relies on matching the fractal codes of the images.

An evaluation of efficiency and effectiveness of indexing techniques for image retrieval is presented in [10]. Other approaches are based on fuzzy searching, taking into account the subjective interpretation of image features [11], and domain specific image distinctive landmarks [12]. In general, databases and retrieval systems designed for specific application fields can use domain knowledge in several forms, in order to improve the classification and retrieval processes.

An approach similar to the one discussed in this paper uses different techniques for feature extraction and image indexing, but is inspired to the same goals [13]. A comparison between the two solutions is forthcoming.

[*] The word *signature* also denotes histograms describing surface contour projections, and is used in machine vision [1]. The two uses of the terms are only related by being both referred to synthetic information extracted from an image.

3. A SIGNATURE FUNCTION BASED ON ANGULAR SPECTRUM

Among the features that describe the visual properties of an image, we study the directionality as a distinguished feature whose indexing can provide effective retrieval in the context of the applications described in Section 1. We argue that orientation of objects within an image is a key attribute in the definition of the similarity with other images, and support this statement with the arguments described in the following of the paper.

In our model of visual interpretation of image content, an image is composed of a background and a foreground. The foreground component is made of one or more areas (objects, parts of objects) whose shape and size are relevant, and define the attention focus. The background surrounds the focus and is often composed of a texture (more or less regular, e.g. grass or clouds in a landscape) or of an almost uniform surface. It is easy to prove the model failure by selecting suitable images, especially in art, advertising, and news fields, but we are assuming that indexing and retrieval do rely neither on the semantic interpretation of the image content, nor on the emotional status of the user.

Therefore the image visual properties are mainly related to the largest foreground components, and among their features the shape, the texture and the orientation play a major role. In many cases shapes can also be defined in terms of presence and distribution of oriented sub-components. For example, a round shape has almost no orientation, i.e., the composing lines are distributed over all the angular range. A thick object has most of its lines arranged along its primary direction, while a slim object has a peak in the line orientation distribution.

Following this assumption, we have defined a metric for image classification based on orientation in the two-dimensional space, that is quantified by signatures composed of angular spectra of image components. It is worth to say that the signature spectra can involve other features like contrast, colours and luminance distribution. The extension will be approached as a continuation of this work.

4. IMAGE ANALYSIS AND SIGNATURE GENERATION

Two major approaches can be followed in extracting information about textures and lines direction in an image. The first is based on images segmented by an edge finding procedure. The second uses full colours or b/w multilevel images.

The edges of the objects in an image can be obtained by computing the convolution of the image (represented as a matrix of pixel intensity values) with a filter matrix. A commonly adopted filter image is the 3x3 matrix

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

The resulting image is classified by a number of directional convolutions and the results are used to compute an angular histogram. Figure 1 shows an example.

By using this technique the image scanning for an angular range [0° - 179°] with small intervals needs a large filter matrix for any distinct angle value. The use of large filter matrix and the scanning for small angular ranges requires an unbearable computational load.

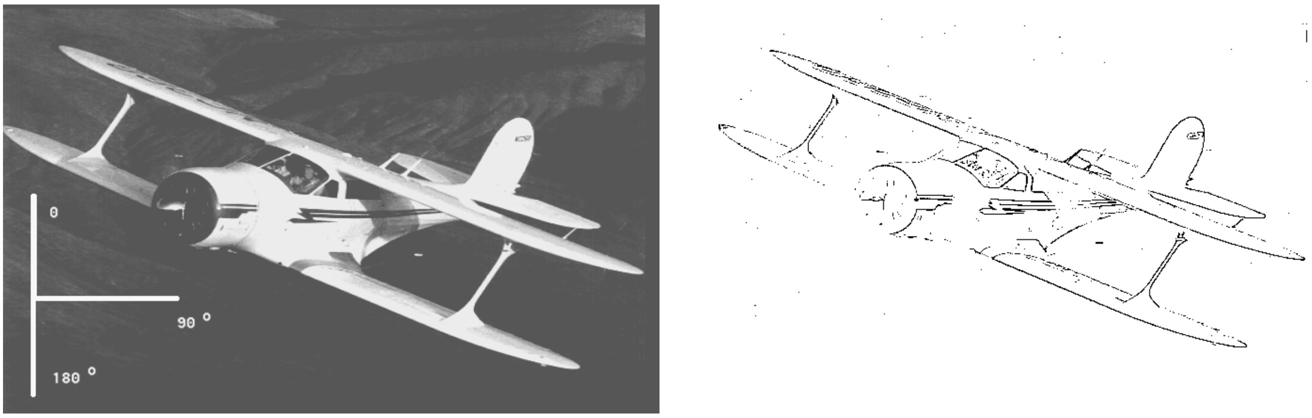
Tamura et al. [8] introduce a fast method for texture analysis based on only two image scans along orthogonal directions. The magnitude $|\Delta G|$ of the image texture is approximated as follows:

$$|\Delta G| = \frac{(|\Delta H| + |\Delta V|)}{2} \qquad q = \tan^{-1}\left(\frac{\Delta V}{\Delta H}\right) + \frac{p}{2}$$

where ΔH and ΔV are the horizontal and vertical components measured by the following 3 by 3 operators :

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

The histogram is computed by quantizing q and counting the points with $|\Delta G|$ value over some threshold t.



(a) (b)

Figure 1. A grey-scale image (a) and the image edges (b)

We approach the problem of finding the distribution of image lines direction by analysing its Fourier transform. As noted by several authors, information direction is preserved in the 2D Fourier power spectrum of an image [14,15].

For a grey-level image, we generate a signature composed of 180 values, one for each angle in [0°-179°] range, each value summing up the size of the image components (i.e., lines) that are shaped along that angle. It is worth to note that we ground our discussion on grey-level images, but the same approach can be applied to colour images, through the luminance component of each pixel value.

In order to compute the Fourier transform, the image needs some pre-processing. We briefly describe the operations performed, since some of them affect the interpretation of the image signal, therefore can be used to focus on specific visual properties.

An anti-aliasing filter is first applied, in order to reduce the high frequencies that constitutes mostly a random noise in the image. A 2D convolution is computed with the following matrix:

$$\begin{bmatrix} 0.25 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

Then, the image is mirrored along the borders in order to obtain a fully symmetric image. This image is four times the original one, and its representation in the frequency domain is characterised by the absence of imaginary components.

The last pre-processing operation is the windowing by a 2D Hamming function, required to transform the 2D signal describing the image into a periodic signal, that can be represented by a finite number of frequency components.

This operation has also an interpretation in terms of image visual properties, since it acts as a low pass filter whose effect is negligible in the center of the image, and increases towards the borders.

The low frequencies are located in large components within the image, which are usually relevant for interpreting the image and can be assumed to be most frequently localised in central areas (e.g., foreground objects) or spanning all the image (e.g., recurrent shapes, landscapes). Their contribution to the image is almost completely preserved.

High frequencies are mostly located in small details and fine-grain textures. Under the assumption stated above, a reduction of these components is equivalent to focusing on the foreground component located in the image central area, and to ignoring the details of the peripheral contour. As we will discuss later, additional filtering of foreground details is accomplished during the analysis and interpretation of the Fourier spectrum.

Once the Fourier transform is computed, its frequency domain representation can be scanned by a revolving vector pointed in the zero frequency component. The vector explores a 180° range, computing the sum of image components contribution for each angle.

In the frequency domain representation continuous and low frequency components are located at the center, while higher frequency components, related to fast variations in the image, are located close to the border.

Most of data is located near the origin. In fact, the image information is largely related to patterns with non-negligible dimension, i.e., large objects, rather than to details and continuous background. We can take advantage of this property and further filter the data by shortening the length *r* of the revolving vector, thus excluding from the computation the high frequency components, related to image details, even if they appear in the focus area of the image.

Similarly, the zero frequency (continuous) component is eliminated since it describes the lowest common grey level, and does not bear useful information. Therefore, the vector that spans the Fourier spectrum is described as $r = [l : h]$, where *l* is the lowest harmonic component and *h* the highest harmonic component considered. Values of *l* and *h* must be found experimentally according to the purpose of the analysis. We tested several values on a sample database of 88 public domain images of planes, and concluded that a vector $r = [1 : 5]$ captures the orientation of the visually most relevant image components.

In order to further reduce noise influence and consideration of components which cannot be regarded as edges of image objects, a threshold *t* is introduced. Components whose value is less that *t* times the maximum value along the same direction are discarded. As for *r*, values of *t* have to be found experimentally. The value $t = 0.15$ has been selected in our tests.

The result of the spectrum analysis is an array of 180 values that represent the total length of image components along the various directions, counted clockwise so that the horizontal direction is at 90°. The values are non-independent for the presence of latching in the 2D spectrum scanning, but this does not influence the significance of the measure.

Figure 2 shows the full angular histogram of the images of Figure 1a (upper diagram) and 1b (lower diagram) traced with parameters $r = [1:20]$ and $t = 0.1$. These values preserve most details of the images, and point out a substantial difference between processing of a grey-scale image and processing of its edges.

The grey-scale image analysis (upper diagram in Figure 2) is characterised by a maximum at 125° corresponding at 35° under the horizontal line clockwise. It is evident that the main visual component in Figure 1a is made by the plane wings at 125°. A second component is represented by the fuselage around 90°.

The image edges analysis (lower diagram in Figure 2) shows a maximum at 90° due to the decoration lines on the fuselage. The contribute of the wings here is lessened by the little number of lines they are composed of.

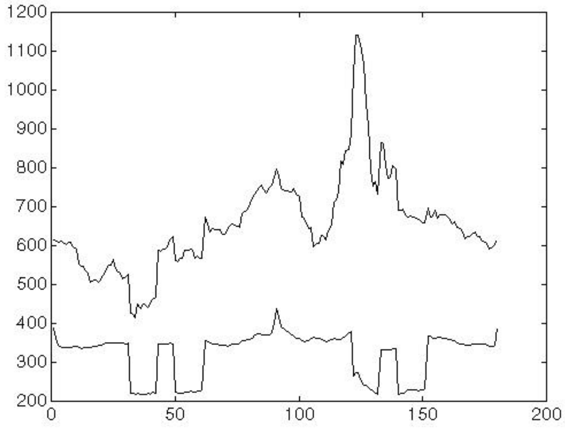


Figure 2. Angular histograms of figure 1 images

5. IMAGE COMPARISON

Retrieval of images by similarity can be grounded on the distance between the image signatures. In our approach the similarity may involve the whole image components, or a specific angular range.

Since the values in the signature represent an integration over the image, the choice of the whole signature seems appropriate when the image exhibits only one main direction. This can easily be confirmed experimentally, and in fact the whole spectrum is used in texture analysis [8]. More frequently, images contain components that visibly suggest a main orientation, together with other components with different layouts. Therefore, using only a subrange of the signature allows more precise retrieval of images that, while presenting a similar overall layout, may differ in details or small components. For example, for the image in Figure 1 a suitable choice for comparison could be the range 90°-140°.

The distance $d_{rif,i}$ between an image *i* and a given reference image *rif* was initially defined as

$$d_{rif,i} = \sum_{r=A,B} \left| q_{rif,r} - q_{i,r} \right|$$

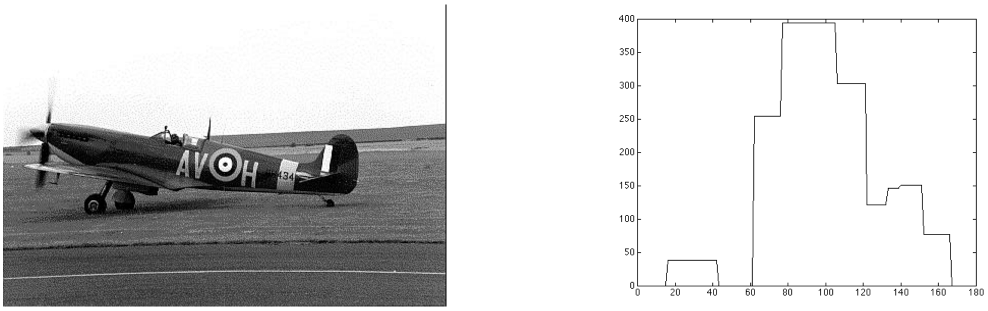
where $q_{i,r}$ is the value of the *i*-th image signature at angle *r*, and *A*, *B* define the angular range.

The distance computed in this way does not consider local differences in the selected range, that mostly appear in the signature as isolated peaks. In order to take into account these differences, we introduce an additional positive term $P_{rif,i}$ if

$$\left| q_{rif,r} - q_{i,r} \right| > K * q_{rif,r}$$

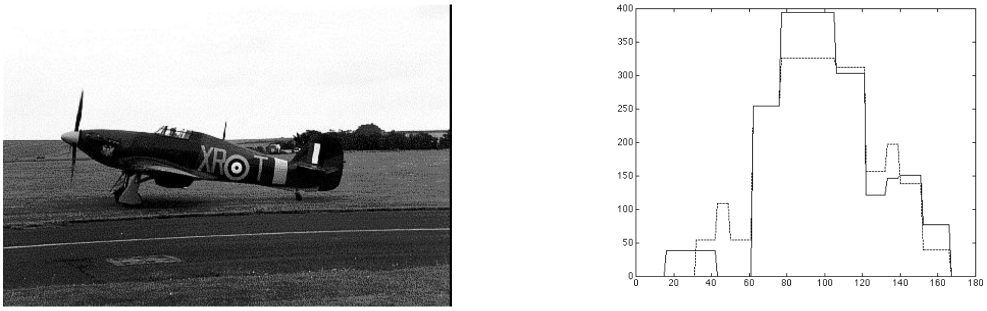
that increases the distance between signatures that are locally almost different, even if they integrate to close values over the selected range. *K* and *P* have to be defined experimentally.

An example of signature match yelding similar images retrieval is given in Figures 3, 4 and 5. Figure 3 shows a reference image and its signature. Spectrum matching is restricted to the range 55°-125° where most of the image content is mapped. In Figure 4 the image with the closest signature is shown.



(a) (b)

Figure 3. The reference image (a) and its signature (b)



(a) (b)

Figure 4. The image closest to Figure 3 (a) and its signature (b) superimposed to signature of Figure 3

Figure 5 shows a set of 9 similar images in rank order, the first one being the reference image. With the exception of image 06, all images are side view of aircrafts on ground, where most image elements are arranged along the horizontal direction. The ground is marked by the horizon line and road borders. Except images 03 and 07, planes are of the same type. Image 03 pictures a plane with a thick fuselage (like the others), and only plane 07 shows a slim shape.

Image 06 is clearly a false drop. The image information is concentrated in small details that have been filtered out with respect to the ground and hangar lines. We shall discuss about it in next section.

6. DISCUSSION

6.1 False drops

Image 06 of Figure 5 is a false drop. In Figure 6 the image and its signature are shown, with the signature of the reference image. In the query range [55°–125°] the two spectra are very close

False drops are generated whenever signatures span a range narrower than the one of the original documents. In our case, being the signature an integral sum of the image contributions along any of the direction, no information is retained about the relative position of the image components, and this may greatly affect the result, especially for small objects.

In this case the false drop comes from the filtering procedure applied during signature generation. In particular, the elimination of the harmonic frequency components over the fifth one has almost deleted the information about the real image focus, that is, the plane crashing on ground in the image right part, that does not appear different from the background.

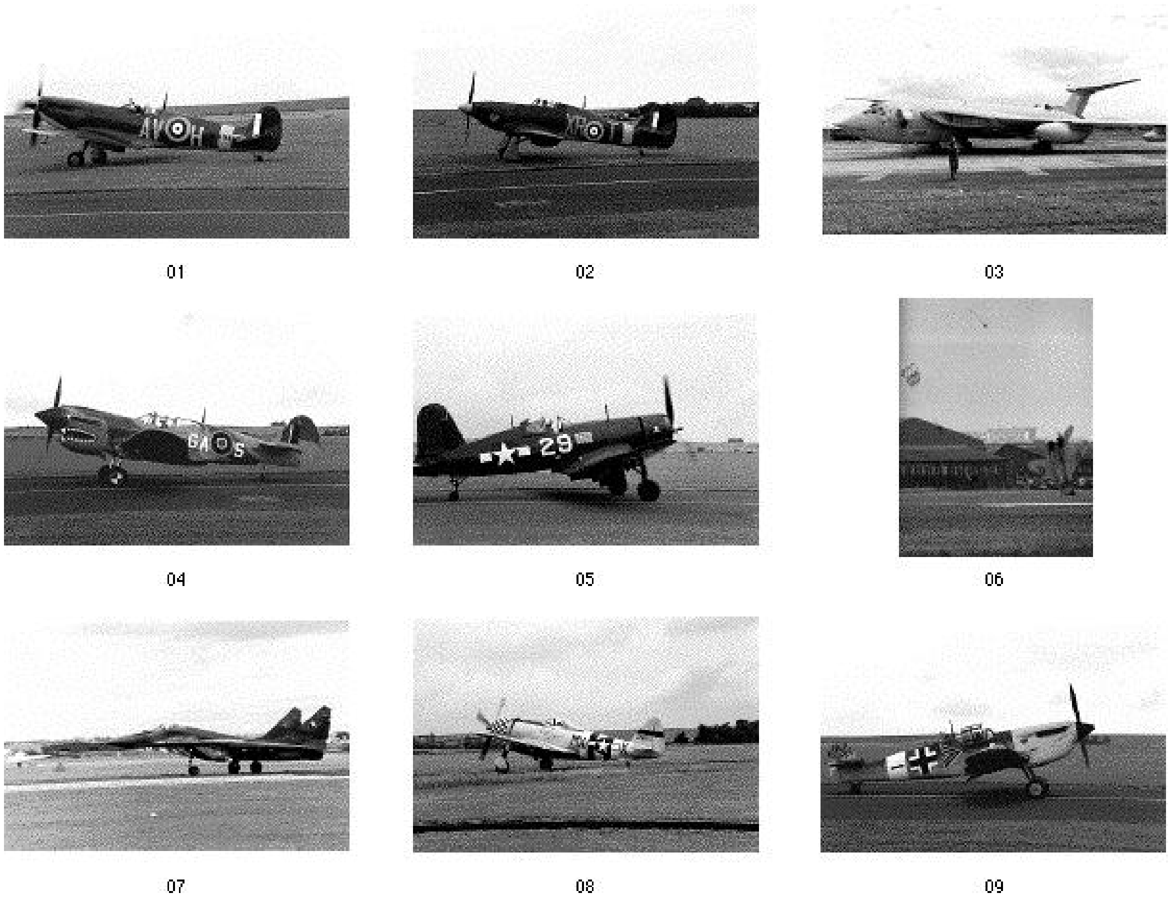


Figure 5. A set of images retrieved by signature matching. First image is the reference image

Finer spectra can avoid such mistakes, but greatly reduce the retrieval recall. A multi-step procedure, starting with coarse-grain spectra and subsequently refining them, could improve the retrieval. We shall approach this problem in the future work.

6.2 Image translation, rotation and mirroring

Image translation, rotation and mirroring may be taken into account by interpreting these operations in terms of the angular spectrum computed on the Fourier transform.

In the frequency domain all the image components are represented as the sum of periodic functions characterised by different periods, centered in the origin (the zero frequency component). Being centered in the origin, they do not retain any information about the original position of the image components, but only information about their direction. This situation is therefore equivalent to the independence of the signature from the image component translation.

Image rotation corresponds obviously to a horizontal shift of the spectrum, and image mirroring corresponds to mirroring of the spectrum.

In all the three cases, however, it is important to note that the discrete nature of the image pixels may introduce noise and aliasing, and that translation and rotations may modify the overall content of the image by adding or excluding some objects.

Noise and aliasing are eliminated by the filters applied before FFT computation. The content modification may require more specific processing.

For example, in a movie the images that belong to the same shot are very similar. Camera movements like panning and tilting are in fact a translation of the image, that does not modify the angular spectrum. But some image components leave the scene, and other components enter. As long as they belong to a uniform background, they may compensate each other. As a foreground object leaves or enters the scene or the background changes, the spectrum variation signals a major modification. Does this mean that the image is now different? The answer cannot be given without considering the real meaning of the objects in the scene.

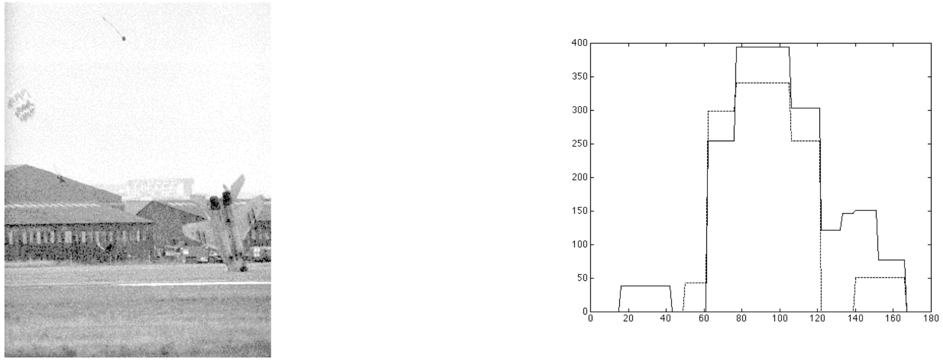
This technique can be confronted with other techniques used for identifying scene boundaries in movies [16,17]. Work on this issue is in progress.

6.3 Performance

The illustrated procedure is based on the FFT computation and its interpretation as an angular spectrum, as we have discussed in Section 4. The computation of FFT is a basic operation in image and signal processing. We have used the Cooley and Tukey formula based on butterfly operation [18].

The total number of floating point operations needed for FFT computation of a N-point sequence is $N \log_2 N$. For a $N \times M$ pixel image, the total number of operations required is therefore

$$M \times N\left(\log_2 M + \log_2 N\right).$$



(a) (b)

Figure 6. A false drop (a) and its signature (b) compared with reference image signature

The complexity of the angular spectrum computation is independent from the image size, and depends only on the number of harmonic terms considered. For a revolving vector $r = [l : h]$ scanning the 180° range the number of floating point operations is

$$180 \times (h - l + 1) + 180$$

the first term referring to the computation of the image contribution along each angle, the second term referring to the normalisation operation. For the examples shown ($r = [1 : 5]$) about 1000 operations are executed. This term is negligible with respect to the FFT computation.

Many programming environments have optimised routines for FFT computation. Some time consumption figures on a Pentium 100 MHz computer running Windows 3.11 are the following ones:

image size	*time (seconds)*
1024*768	159
640*480	56
320*240	13

7. CONCLUSION

In this paper we have discussed the use of the angular spectrum of an image as a signature for comparing image content. Content based retrieval can be based on spectrum comparison between a reference image, assumed as the query, and an image collection. In order to be effective, the collection must be semantically homogeneous so that visual similarity can be a surrogate for content match.

We have introduced a technique for computing the image angular spectrum that is based on the Fourier transform. The technique is not based on image annotation, therefore it does not require human assistance. Efficient routines are available in almost all programming environment, that allow the database population process to be executed in acceptable time.

We have also introduced a metric for comparing the spectra and ranking the result, and approached the issue of partial query specification. Sample results on a small test collection have been given, and the problem of false drops has been discusses.

The approach is suitable for image retrieval application where the graphical content or layout bear most of the information. Initial tests in specific domain like the analysis of human signatures and fingerprints, are promising. However they have to be compared with the safer and more assessed techniques that are used in this cases.

The use of direction based signatures seems also suitable for scene boundary detection and frame sorting in movies.

8. ACKNOWLEDGMENT

This work has been supported by MURST, Italy, in the framework of the project *“Basi di dati evolute: modelli, metodi e strumenti”*.

9. REFERENCES

1. R. M. Haralick, L. G. Shapiro, *Computer and Robot Vision*, Addison-Wesley, 1992
2. *ACM Multimedia Systems*, special issue on Multimedia Databases, 3, 5/6, 1995
3. Barber et al., “Query by Content for Large On-line Image Collections”, *Multimedia Systems, an IEEE Tutorial*, IEEE Press, 1994.
4. *IEEE Computer*, special issue on Content Based Image Retrieval, 28, 9, 1995
5. M. Flickner et al., “Query by Image and Video Content: The QBIC System”, *IEEE Computer*, 28, 9, 1995

6. V. E. Ogle, M. Stonebraker, “Chabot: Retrieval from a Relational Database of Images”, *IEEE Computer*, 28, 9, 1995

7. K. P. Qing, W. R. Caid, C. Ren, P. McCabe, “Image/Text Automatic Indexing and Retrieval System Using Context Vector Approach”, in C. C. Jay Kuo (ed.), *Digital Image Storage and Archiving Systems, SPIE 95 Proceedings*, Philadelphia, October 25-26, 1995

8. H.Tamura, S.Mori, T. Yamawaki, “Textural Features Corresponding to Visual Perception”, *IEEE Trans. on Systems, Man and Cybernetics*, 8, 6, 1978

9. A. Zhang, B. Cheng, R. Acharya, “An Approach to Query-by-Texture in Image Database Systems”, in C. C. Jay Kuo (ed.), *Digital Image Storage and Archiving Systems, SPIE 95 Proceedings*, Philadelphia, October 25-26, 1995

10. J. Barros, “Trading Efficiency for Effectiveness in Similarity-Based Indexing for Image Databases”, in C. C. Jay Kuo (ed.), *Digital Image Storage and Archiving Systems, SPIE 95 Proceedings*, Philadelphia, October 25-26, 1995

11. J. K. Wu, A. S. Narisimhalu, “Identifying Faces Using Multiple Retrievals”, *IEEE Multimedia*, 1, 2, 1994

12. J. Bach, S. Paul, R. Jain, “A Visual Information Management System for The Interactive Retrieval of Faces”, *IEEE Trans. on Knowledge and Data Engineering*, 5, 4, 1993

13. E. Di Sciascio, A. Celentano, “Similarity Evaluation in Image Retrieval Using Simple Features”, in *Storage and Retrieval for Image and Video Databases V, SPIE's Electronic Imaging '97*, San José, February 8-14, 1997

14. S. K. Mitra, J. F. Kaiser, *Handbook for Digital Signal Processing*, John Wiley & Sons, 1993

15. A. V. Oppenheim, R. W. Schafer, *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, 1975.

16. H. Zhang, A. Kankanhalli and S.W. Smoliar, “Automatic Partitioning of Full Motion Video”, *ACM Multimedia Systems*, 1, 1, 1993.

17. S.W. Smoliar, H. Zhang, “Content-Based Video Indexing and Retrieval”, *IEEE Multimedia*, 1, 2, 1994

18. J.W.Cooley, J.W.Tukey, “An algorithm for the machine calculation of complex Fourier series”, *Math.Comp.* 19, 1965.