About optimal use of color points of interest for content-based image retrieval

Valérie Gouet, Nozha Boujemaa

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

About optimal use of color points of interest for content-based image retrieval

Valérie Gouet — Nozha Boujemaa

N° 4439

Avril 2002

THÈME 3

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4439--FR+ENG

About optimal use of color points of interest for content-based image retrieval

Valérie Gouet , Nozha Boujemaa

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Imedia

Rapport de recherche n° 4439 — Avril 2002 — 23 pages

Abstract: In content-based image retrieval systems, the main approaches based on the query-by-example paradigm involve an approximate search on the whole image, which requires a global description of it. When considering tasks like object recognition or partial queries on particular area, these methods become inadequate and more local characterizations must be employed. In this context, image description based on points of interest appear best adapted. The point characterization which proved reliable is based on invariants to rotation and in particular on combinations of the Hilbert's differential invariants. For gray value images, such a description used to be considered up to third order at least. More recently, generalizations to color images were proposed for stereovision and image retrieval. Some of them propose to consider the invariants only at first order and to enrich the characterization with geometrical constraints for describing spatial relations between points, while others consider higher order invariants and compute some combinations of them to achieve illumination changes invariance. In this report, we discuss the advantages and drawbacks of these different choices, with the aim of proposing an optimal use of color points of interest for content-based image indexing and retrieval.

Key-words: Image indexing and retrieval, Partial queries, Color images, Local descriptors, Differential invariants.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Utilisation optimale des points d'intérêt couleur pour la recherche d'images par le contenu visuel

Résumé : En indexation d'images par le contenu visuel, les principales approches basées sur le paradigme de la recherche par l'exemple impliquent une recherche approximative sur la totalité de l'image. Ce type de requête passe par une description globale de l'image. Lorsque l'on s'intéresse à la reconnaissance d'objets ou aux requêtes partielles sur une zone particulière, ces techniques deviennent inadéquates et des descripteurs plus locaux doivent être envisagés. Dans ce contexte, la caractérisation de l'image à base de points d'intérêt s'est révélée la plus adaptée. La description de points alors usuellement associée est basée sur un ensemble d'invariants à la rotation et plus précisément sur une combinaison des invariants différentiels de Hilbert. Pour les images en niveau de gris, une telle description doit être considérée au moins jusqu'à l'ordre 3. Récemment, des généralisations à la couleur ont été proposées pour la stéréovision et la recherche d'images. Certaines d'entre elles proposent de considérer les invariants uniquement à l'ordre 1 et d'enrichir la caractérisation par des contraintes géométriques décrivant les relations spatiales entre points d'intérêt, alors que d'autres impliquent les invariants à des ordres supérieurs et en calculent une combinaison visant à obtenir la robustesse aux changements d'illumination. Dans ce rapport, nous étudions les avantages et inconvénients de ces différents choix, dans le but de proposer une utilisation optimale des points d'intérêt couleur pour l'indexation et la recherche d'images par le contenu.

Mots-clés : Indexation et Recherche d'images, Requêtes partielles, Images couleur, Descripteurs locaux, Invariants différentiels.

Contents

1 Introduction

In content-based image retrieval (CBIR) systems, the query-by-example paradigm usually involves the entire image. Most often, global properties of image visual contents are extracted. On the other hand, user could be interested to retrieve only similar parts or objects of images. In this context, global image descriptors are unusable. For this purpose, a more local solution must be considered, where each region of the signal is analysed independently. More precisely, the concept consists first in localizing relevant regions in the image and second in characterizing the primitives obtained, by considering local descriptors. Seeing that no information about the image database contents is known a priori, we are looking for techniques which are stable with respect to possible image perturbations like partial occultation, viewpoint or illumination changes.

In this context, mainly two classes of primitives are relevant to characterize part or object in an image: region [18, 2, 5] and points of interest. In this paper, we are focused on points of interest for some reasons listed below:

First, points of interest extraction performs well whatever the image contents is, unlike region-based approaches in which the quality of the segmentation step is sensitive to image geometrical contents. Moreover, points are more robust to geometric transformations of the image like viewpoint changes, since the description is computed locally, and also robust to partial occultation, unlike other geometric primitives of higher level. Finally, content-based image retrieval techniques exploit photometric information contained in the images. By definition, a point of interest is located where this photometric information is the most significant. Then using points, we have the right to expect a local image characterization rich and compact at once.

Matching points of interest is a classical task in Computer Vision for applications like stereo rig calibration or 3D reconstruction, see for example [19, 9]. Since 1996 with the works of Schmid on point extraction and characterization, these primitives proved their utility for content-based image retrieval applications. However, it is important to consider that their conditions of use are different of the stereovision ones, and it is not possible to directly transpose to image retrieval the solutions developed for stereovision. For example, the growing volume of the image databases do not allow to consider all the geometrical constraints available or which can be estimated between two images, like the epipolar constraint. It implies too the need of computing the most compact possible indexes from the photometric information. Another argument concerns the quality of the collected images, which mostly cannot be controlled. The image characterization should be robust to image coding and possible destructive compression. In this context, we studied in literature some alternatives describing the way to characterize images from points of interest. But none of the solutions presented was compared to the others in practical cases. Dealing with point characterization using differential invariants, we encountered several alternatives involving feature vectors of very different sizes. The purpose of this paper is to discuss the advantages and drawbacks of these different choices, with the aim of proposing an optimal use of color points of interest for image retrieval.

The paper is structured as follows. Section 2 consists in a brief state of art of the existing image characterization techniques based on color interest points and differential invariants, followed by a presentation of the possible strategies for indexing and doing retrieval on them. To compare the different solutions, we present in section 3 a study on the optimal way to use the color point-based characterization, by evaluating the impact of different photometric and geometric features on the indexing and retrieval procedure. The results obtained are discussed in the conclusion.

2 Image indexing and retrieval using points of interest

Image indexing based on points of interest consists first in being able to find informative points in each image of the database. In section 2.1 we revisit point extraction and characterization using differential invariants particularly, before considering in section 2.2 the image retrieval strategy to apply on databases indexed by points of interest.

2.1 Image characterization using points of interest

When applied to image retrieval, image matching based on points of interest needs points with excellent *repeatability*, ie. points that can be extracted from an image to another with the same accuracy and under various conditions like viewpoint or illumination changes. Many point extractors exist for gray value images, but it is a color operator which fits better that required repeatability: the Harris color detector [16, 8]. We revisit it in the next section.

2.1.1 Harris color points of interest

The Harris color extractor computation only implies the first order derivatives of the image. Its formulation depends on the following matrix M:

$$M = G(\widetilde{\sigma}) \otimes \begin{bmatrix} R_x^2 + G_x^2 + B_x^2 & R_xR_y + G_xG_y + B_xB_y \\ R_xR_y + G_xG_y + B_xB_y & R_y^2 + G_y^2 + B_y^2 \end{bmatrix} \tag{1}$$

where $G(\widetilde{\sigma})$ stands for a gaussian smoothing of size $\widetilde{\sigma}$ and (I_x, I_y) represents the first derivatives of the channel I with $I \in \{R, G, B\}$ numerically implemented using a gaussian convolution of size σ. The color points of interest are defined by the positive local extrema of the following operator :

$$\mathrm{Det}(M) - k\,\mathrm{Trace}^2(M) \quad \text{with} \quad k = 0.04$$

After this thresholding, the points associated to the n best local extrema are kept.

An adaptation robust to scale changes is possible and has been proposed in [15] for the gray value version of the Harris extractor. The method is robust up to a scale factor of 4 ; it consists in computing a multi-scale representation for the detector and then in selecting points at which a local measure (the Laplacian) is maximal over scales.

In the next section, we revisit the approaches based on differential invariants to characterize the Harris color points.

2.1.2 Characterization using color differential invariants

In [17], Schmid proposed to describe an image by using the Harris points of interest characterized with photometric local descriptors computed from gray value images. Such a description involves the Hilbert's invariants [11] and more precisely some combinations of them in order to get translational as rotational invariance, and to ensure numerical stability. These quantities can be computed at order 1, 2 or 3 (higher order are increasingly complicated). See equation 2 below for their formulation up to third order at a given point:

$$\left\{ \begin{array}{l} I \\ I_x^2 + I_y^2 \\ I_{xx}I_x^2 + 2I_{xy}I_xI_y + I_{yy}I_y^2 \\ I_{xx} + I_{yy} \\ I_{xx}^2 + 2I_{xy}^2 + I_{yy}^2 \\ I_{xxx}I_y^3 + 3I_{xyy}I_x^2I_y - 3I_{xxy}I_xI_y^2 - I_{yyy}I_x^3 \\ I_{xxx}I_xI_y^2 + I_{xxy}(I_y^3 - 2I_x^2I_y) + I_{xyy}(I_x^3 - 2I_xI_y^2) + I_{yyy}I_x^2I_y \\ I_{xxy}(2I_xI_y^2 - I_x^3) + I_{xyy}(I_y^3 - 2I_x^2I_y) - I_{yyy}I_xI_y^2 + I_{xxx}I_xI_xI_y \\ I_{xxx}I_x^3 + 3I_{xxy}I_x^2I_y + 3I_{xyy}I_xI_y^2 + I_{yyy}I_y^3 \end{array} \right. \tag{2}$$

Generalization to color images was introduced using first order invariants for stereoscopic images matching in [16], for sub-image retrieval in [6] and up to third order in [10] for image matching under illumination changes. The color description presented at first order is based only on the first order invariants of each channel and enriches them with two first order quantities specific to color, to take in the following set:

$$\nabla R.\nabla G \qquad \nabla R.\nabla B \qquad \nabla G.\nabla B \tag{3}$$

Such a description contains 8 invariants of first order, listed at equation 4 below:

$$\vec{v}_{col}(\vec{x}, \sigma) = \begin{pmatrix} R \\ \|\nabla R\|^2 \\ G \\ \|\nabla G\|^2 \\ B \\ \|\nabla B\|^2 \\ \nabla R.\nabla G \\ \nabla R.\nabla B \end{pmatrix} \tag{4}$$

It is almost equivalent to the gray value implementation in terms of storage cost (8 invariants instead of 9). It has been proved to be more robust because of the sensitivity to noise of the second and third order derivatives, which may be worsened by the numerical complexity of the corresponding invariants computation, as established in [12]. On the contrary, the color description implies first order magnitudes particularly stable at Harris color points location, since by definition these points are extracted at local extrema of the first order derivatives of the signal (cf. section 2.1.1).

2.1.3 Considering other image transformations

This section mainly deals with invariance to illumination changes, that is the so-called problem of *color constancy* in content-based image indexing and retrieval applications. In addition, we briefly revisit the multi-scale implementation of the feature vectors for invariance to scale and robustness to changes of viewpoint by considering affine invariants.

Color constancy

Image transformations must include the illumination changes, which represents a very classical problem in image indexing tasks and has generated many works, mainly for global image description. The case of invariance to this class of changes requires first to choose an illumination model. Recently, local description aspects has been considered in [10], where eight models are evaluated for color images. Experiments show that the diagonal model with translation (DT), having 6 parameters, is characterized by the best ratio quality/complexity for local context, what is the case with differential invariants. Its formulation is presented at equation 5 below:

$$\bar{I(\vec{x})} = D.I(\vec{x}) + \vec{T} \tag{5}$$

where D is a (3×3) diagonal matrix and $\vec{T}$ a translational vector of dimension 3.

On the basis of the DT model, two alternatives can be considered to achieve photometric invariance:

- *Normalization of the local descriptors.* Considering ratios of low order invariants allows to eliminate all the parameters of affine illumination models. For instance, by introducing ratios of the gradient magnitude, the normalization of the gray value invariants of equation 2 provide the set of invariants of equation 6:

$$\left\{ \begin{array}{l} \frac{I_{xx}I_x^2+2I_{xy}I_xI_y+I_{yy}I_y^2}{(I_x^2+I_y^2)^{\frac{3}{2}}} \\ \frac{I_{xx}+I_{yy}}{(I_x^2+I_y^2)^{\frac{1}{2}}} \\ \frac{I_{xx}^2+2I_{xy}^2+I_{yy}^2}{I_x^2+I_y^2} \\ \frac{I_{xxx}I_y^3+3I_{xyy}I_x^2I_y-3I_{xxy}I_xI_y^2-I_{yyy}I_x^3}{(I_x^2+I_y^2)^2} \\ \frac{I_{xxx}I_xI_y^2+I_{xxy}(I_y^3-2I_x^2I_y)+I_{xyy}(I_x^3-2I_xI_y^2)+I_{yyy}I_x^2I_y}{(I_x^2+I_y^2)^2} \\ \frac{I_{xxy}(2I_xI_y^2-I_x^3)+I_{xyy}(I_y^3-2I_x^2I_y)-I_{yyy}I_xI_y^2+I_{xxx}I_xI_xI_y}{(I_x^2+I_y^2)^2} \\ \frac{I_{xxx}I_x^3+3I_{xxy}I_x^2I_y+3I_{xyy}I_xI_y^2+I_{yyy}I_y^3}{(I_x^2+I_y^2)^2} \end{array} \right. \tag{6}$$

This approach is reasonably possible only for invariants of high orders ; indeed at first order, no invariant would remain for gray value images and the set would be dramatically reduced to $8-6=2$ invariants for color images. This solution is applied in [10] for color image matching by using $29-6=23$ invariants up to third order.

- *Image normalization.* In practice, introducing ratios of derivatives in the characterization as evoked below may worsen the unstability of the invariants against noise, by making the formulas even more sophisticated. In this context, another alternative consists in normalizing the image with respect to the illumination model, before computing the invariants. This approach has been studied in [7] for the color invariants of first order. A local affine normalization of the image gray values is proposed for approximate internal and external illumination changes, according to the DT model. The experiments show that the 8 invariants gain in robustness, when computed after image normalization.

The table 1 recapitulates the number of differential invariants for gray value and color images up to order 3. The second row presents the number of quantities obtained when the gray value invariants are normalized according to an affine illumination model (2 parameters) ; the last one presents the number of color invariants when the DT model is considered for illumination changes normalization.

	Order 1	+ Order 2	+ Order 3
Gray value differential invariants	2	5	9
Invariance to affine illumination changes (2 ddl.)	0	3	7
Differential invariants specific to color	2	2	2
Total for color images	8	17	29
Invariance to illumination changes (DT - 6 ddl.)	2	11	23

Table 1: Gray value/color differential invariants enumeration.

Multi-scale approach

For important camera shifting or focal changes, a multi-scale approach is required to compute the invariants. An approach robust up to scale factor of 4 is proposed in [15] by computing the descriptors for each point of interest at its characteristic scale (which has been determinated during the point extraction step). Here the descriptors used are steerable filters computed up to order 4, but the same approach can be exploited for Hilbert's differential invariants.

Viewpoint changes

Adding the invariance property just presented to translational and rotational invariance, the characterization becomes invariant to the similitude group, then quasi-invariant to perspective transformations [3], and consequently robust to small viewpoint changes.

For unconstrained changes of viewpoint, a small planar surface patch undergoes an affine transformation in the images. Consequently, recent works deal with affine invariants computation. It is possible to extend the idea of searching points in the Gaussian scale space representation to affine Gaussian scale space. The approach has been introduced in [13] where an affine invariant neighborhood is estimated iteratively. For our purpose, the point extraction consists in computing the 2nd moment matrix involved in the operator of equation 1 in an ellipse instead of a circular window. Then for each point of interest, the descriptors are computed from a normalized patch around the point for which stretch and skew have been removed. This approach is used in [1] for uncalibrated wide baseline stereo by considering Fourier-Mellin descriptors on color images.

2.1.4 Similarity measure

Let $\mathcal{V}$ be the feature space of the color points of interest according to the local characterization just introduced. Because each invariant has different magnitudes, $\mathcal{V}$ is an hyperellipsoide. The most suitable metrics related to such feature spaces remains the Mahalanobis distance δ^2, because it uses the covariance matrix of the point components and thus takes into account the difference of their magnitude, but especially because this matrix integrates a model of noise of the components ; it represents a crucial aspect especially for second and third orders invariants, which are considered to be more noisy. Traditionnally, the covariance matrix is estimated statistically over the whole points database, or better over a large set of representative points samples. Another interesting advantage is that δ^2 is χ^2 distributed, so that matches associated to the worst distances can be eliminated by thresholding. The Mahalanobis distance is more complex in term of time computation than other distances, but a singular decomposition of the covariance matrix allows to apply on the components a change of basis in order to transform the hyperellipsoide $(\mathcal{V}, \delta^2)$ into a hypersphere, before using the $L2$ metrics.

2.2 Indexing and searching strategy

If we consider the local descriptors presented above, an image is represented by a set of n points $\{p^i\}$ characterized in a feature space $(\mathcal{V}, \delta^2)$ whose size depends on the orders 1, 2 or 3 considered for the invariants computation, and where δ^2 represents the Mahalanobis distance. In this context, building an index for a database $\{I_j\}$ of N images consists in computing a set of $n \times N$ feature points in $(\mathcal{V}, \delta^2)$. Searching for an image or part of an image in the indexed image database consists in *range-queries*, which comes to find in this space the closest points to the query points.

Let $\{q^i\}$ be the set of query points. As illustrated in figure 1 by considering the color description of the points of interest introduced at equation 4, the retrieval strategy consists in a voting algorithm. The $\{c^i\}$ closest points to the $\{q^i\}$ query points are characterized by scores which are function of the distance between c^i and q^i. A vote is computed for each image, by considering a combination of the scores related to the matches (c^i, q^i) involved in the image. The images most similar to the query are characterized by the best votes. The complexity of the query can be efficiently reduced by arranging the indexes according to multidimensional structures.

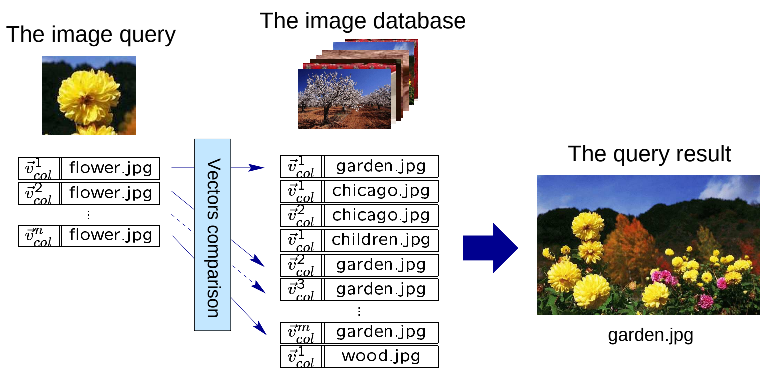


Figure 1: The indexing and searching strategy.

Such a characterization can be enriched by considering spatial relations between points of the same image. This class of geometrical constraints are semi-local since they are based on the neighborhood of the points to match. We present in the next section the constraints encountered in literature.

Neighborhood constraints

The basic idea usually encountered consists in considering a couple of 2D points (m_1, m_2) as a potential match if in the neighborhood of m_1 there are enough points matched with neighbors of m_2.

Some heuristics based on geometrical constraints can be added to precise the spatial configuration of the neighbors. It depends on the transformations existing between the images to match. The characterization approach based on differential invariants was defined to be invariant to similitude group. Then we need to define constraints which take into account these degrees of freedom.

A solution consists in considering the property of angles conservation between neighbors ; it supposes that the rigidity constraint is respected. Some authors [17, 14] propose to consider the angle defined between two neighbors of the point to match, which must be roughly constant from an image to another. In [6], an angular constraint implying a complexity of minor importance is defined between the point considered and only one of its neighbors.

The angle between their respective gradients must be preserved from an image to another, as illustrated in figure 2.

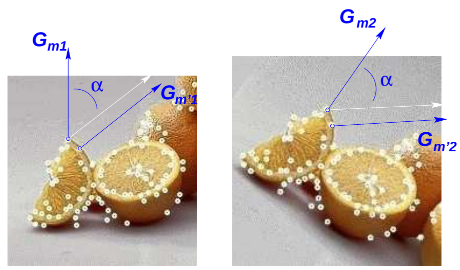


Figure 2: Conservation of the angles between neighbors gradients.

Here the gradient used for the geometrical constraint is the multispectral gradient of Di-Zenzo [4] computed directly from the RGB space. It is assumed to be more robust to noise, since it implies all the channels of the image simultaneously.

We revisited in this section the way to index and retrieve images by using points of interest chararacterized with differential invariants. We saw that some alternatives are possible, concerning in particular the choice of the order for invariants computation, or the use of geometrical constraints. These different alternatives are encountered in the literature but nowhere it was made a comparison for an optimal use of the point-based characterization. Such a study is the subject of the following section.

3 About optimal use of color points of interest

In order to get optimal indexes for image retrieval application, we try in this section to answer many questions still not solved about the concrete use of color differential invariants for image matching. Seeing that authors seem to implement them at different orders, a first experiment consisted in section 3.1 in determining if high orders have a real impact for color description. Given that in practice the image acquisition process cannot be always controlled, section 3.2 is devoted to test their robustness against image coding, like the classical JPEG one. Their adaptation to illumination changes is studied in 3.3. Finally, the contribution of the gometrical constraints on neighborhood is concretely evaluated in section 3.4.

Benchmark images

The image database used for most of the experiments is a subset of the Columbia color image database *coil-100*. We considered here 1000 images representing 50 rigid objects taken under 20 points of view corresponding to a 3D rotation up to 105 degrees. Images related to higher angles were not kept in order to ensure correct matching rates for each image.

This database does not deal with scale changes. Indeed here the aim is not to experiment until which point the color characterization is robust to scale or viewpoint changes, but to compare its relevance while modifying particular intrinsic and extrinsic parameters. Some of the images used are presented in figure 3. For the experiments on illumination changes of section 3.3, a particular database will be considered.

Figure 3: Samples of the Columbia color objects database.

This benchmark database was indexed using the Harris color points of interest extractor. Each point was characterized with the class of color differential invariants of equation 4 and computed in $(\mathcal{V}, \delta)$ at order 1, 2 or 3 according to the experiments. Due to the particular set of images used, the affine multi-scale approach was not implemented here. Approximately 30 points were extracted per image, leading to a database index of 30000 feature points. No geometrical constraints were used, except for the specifical experiments of the last section.

Precision/recall diagram

Most of the results in this section are presented as precision/recall diagrams. If we consider N classes of N_C elements (here $N = 50$ and $N_C = 20$), the precision for each image query is the proportion of retrieved images belonging to the query class; it is computed on the r first retrieved images, where r is the recall parameter ($1 \leq r \leq N_C$). The graphs present the average precision obtained for all the queries, i.e. for all the images of the ground-truth database.

Now let us consider some aspects of the point of interest characterization which look interesting to examine.

3.1 Which order for the color differential invariants ?

Experiments in [16] has shown that the use of color information leads to a point description richer than using gray value invariants, even if only implemented at first order (see section 2.1.2). In this context, it would be now interesting to consider higher orders for color point description. Then we have implemented the color differential invariants up to third order and indexed the benchmark database by considering them at these different orders. The precision/recall diagrams obtained are presented in figure 4.

Roughly speaking, we see that the average precision varies from 0.84 to 1 for all the merged orders, i.e. that at least the four-fifth of the N_C first images retrieved belong to the query class. Of course all the query images were returned for $r = 1$. We noticed visually for some particular objects that the images belonging to the query class and retrieved at first

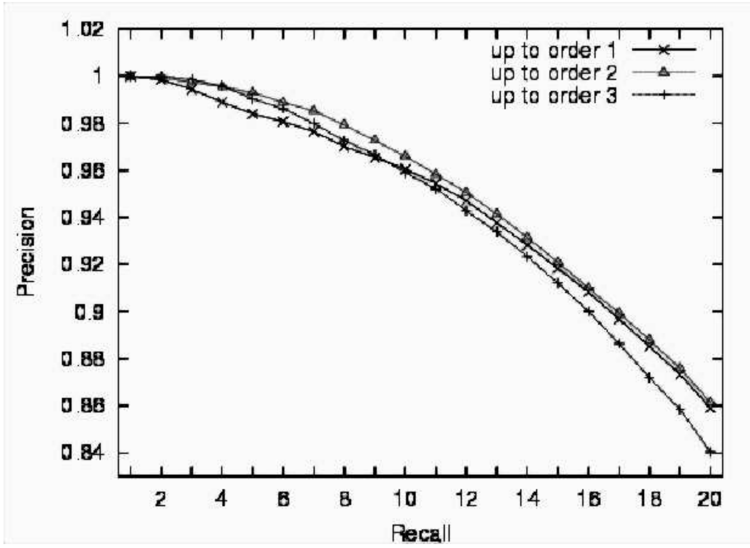


Figure 4: Precision/recall diagram according to different invariants orders.

correspond to the smallest changes of viewpoint, whereas the precision parameter decreases inversely with the angle of view. These considerations are not surprising since it illustrates the fact that the differential invariants are quasi-invariants and then robust to viewpoint changes up to a certain point.

With regard to the relevance of the invariants used at order 2 and 3 in comparison to order 1, the graphs show that employing order 2 slightly improves the responses for all the viewpoints. For $r = 8$, the results are 1% better, which means that order 2 allows to find an average of 0.08 images more in the 8 first images returned. It is especially relevant for the first part of the images retrieved ($r \leq \frac{N_C}{2}$), while becoming hardly perceptible for higher viewpoints. At order 3, the results are a bit better than for second order only for the third first images returned but decrease for more different points of view, until to become the worst for the second part of the retrieved images (with a difference of 2%). These results clearly show that the color characterizations based on third order (and on second order in a lesser extent) are not stable faced with changes of viewpoint. This may appear disappointing since 29 quantities instead of 8 have been employed at order 3, but it can be explained by the following considerations: First the instability of the derivatives increases with order, according to the well-known idea that any measurement of variation amplifies the noise. Second, the formulation of the invariants from partial derivatives becomes more and more complex with order (see equation 2) and consequently amplifies the noise produced by high order derivatives even more.

These remarks should not be ignored, because there are many possible noise sources for local descriptors: one should consider for instance noise related to image acquisition, noise produced by image coding and possible destructive compression. Without forgetting that points extractors suffer from delocalization, in spite of the fact that the Harris color one is characterized by a good repeatability. This problem contributes to increase the instability of high order invariants computed at these points ; in parallel, the first order ones are less disturbed by this inaccuracy, since their computation are based on first order derivatives corresponding to local maxima at these points, by definition of the Harris color operator.

In the next section, we look further into the concept of sensibility to noise of the local descriptors by considering the JPEG coding case.

3.2 Local descriptors robustness and Jpeg coding

The images used for the previous experiments are not free of coding noise, since the ground-truth database was converted from PPM into JPEG format, which is a coding mode that generates losses. However, we made sure to minimize this noise by using a compression quality of 100%. Unfortunately in practice, one cannot rely on the quality of the collected images which can be of any kind. In this context, it is important to consider the coding noise in the evaluation of the local color descriptors for image indexing and retrieval.

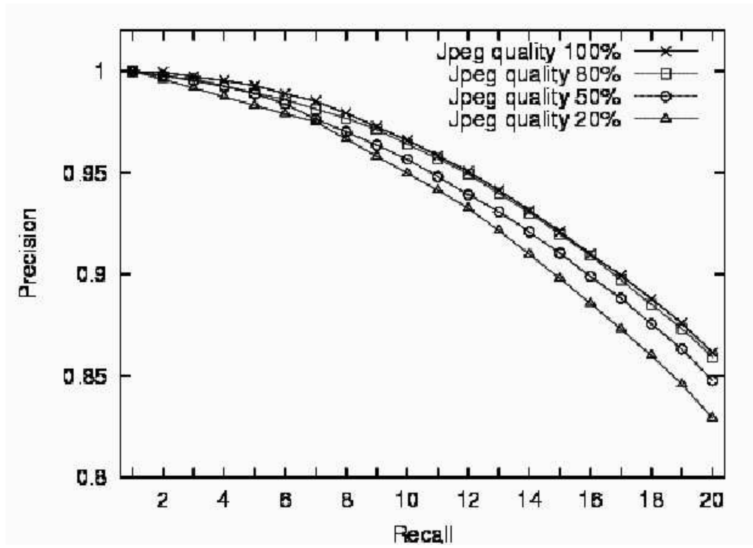
In [12], Jolion made experiments on the influence of JPEG coding on gray value image descriptors and particularly on local descriptors by observing the stability of derivatives, points of interest and differential invariants for different compression rates. He estimates for each of the studied descriptors the minimal threshold of compression required for considering the descriptor sufficiently robust to coding noise. The minimal quality expected is 65% for first order, 80% for second order and 95% for third one. These results do not take the detector delocalization phenomenon into account, since the invariants have been computed on points extracted before image compression. Dealing with gray value points extraction, the minimal quality recommended to extract points with quite stable location is about 70%.

According to these considerations, we have experimented the problems due to JPEG coding on color images indexing. Three databases were generated from the ground-truth database, setting respectively JPEG quality rates to 80%, 50% and 20%. They were then indexed using the color differential invariants up to third order. The precision/recall curves obtained for second and third orders are presented in figure 5 with the ones related to the not compressed images and computed at previous section. The curves related to first order only are not proposed here because they remain very similar to each others, in spite of the compression rate applied.

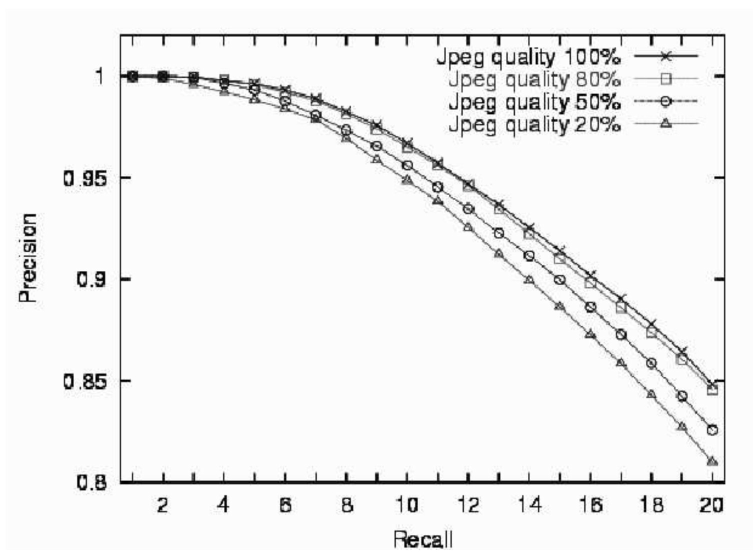
We see clearly the influence of image compression for second and third orders, whereas the precision remains roughly identical at first order whatever the rate applied. These results are in agreement with the conclusions of Jolion concerning the sensibility to noise and point delocalization of high orders differential quantities.

Figure 6 gathers the precision/recall diagrams obtained for orders 1, 2 and 3 after the JPEG coding keeping a quality of 50%. This average compression rate is not unrealistic in practice where images may be compressed several times.

The curves obtained traduce that in practical cases, it is impossible to affirm that using second order enrich the characterization clearly, since matching results become worse than



Orders 1,2



Orders 1,2,3

Figure 5: Influence of JPEG coding on image indexing using the color invariants.

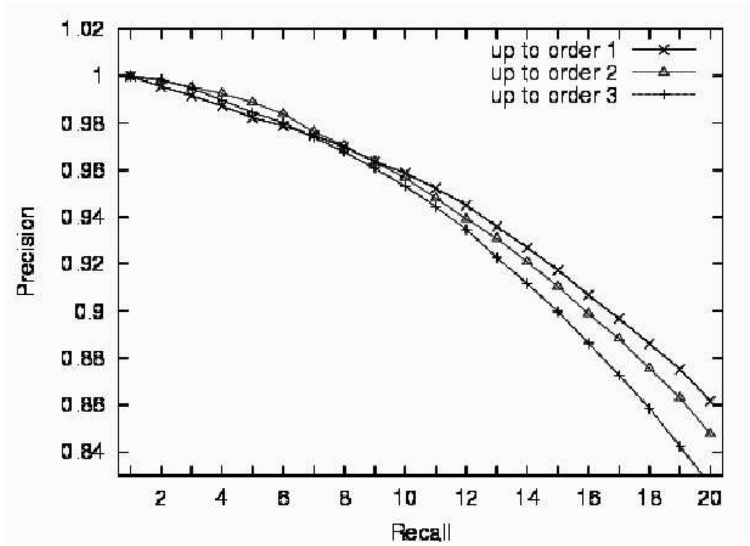


Figure 6: Precision/recall diagram obtained after a JPEG compression keeping a quality of 50%.

using only order 1 for important viewpoints. On the other hand, one can reasonably say that third order characterization is to be proscribed for the class of images used.

From these considerations, we should be careful concerning the use of high orders invariants, which produce disappointing results as regards the number of invariants involved. At this stage of the study, our opinion is that, when the image acquisition cannot be controlled, it is reasonable to consider invariants up to second order only. The third order characterization should not be proscribed, indeed it makes sense for particular applications involving high quality images differing from very small viewpoints, like stereoscopic satellite images.

3.3 About color constancy

At section 2.1.3, we presented two ways for gaining invariance to illumination changes, based for the first on the normalization of the invariants, and on the normalization of the image for the second one. According to the conclusions of the previous section, the characterizations that we are going to experiment for the color constancy problem are the ones involving first and second orders. The third order invariants were proved to be noisy, and this disadvantage would be dramatically accentuated by mixing them to eliminate the parameters of the illumination model.

We consider for this experiment an image ground-truth database of 20 tridimensional objects represented under 11 different illuminants and some small viewpoint changes[1]. An example is presented in figure 7.

Figure 7: One of the 3D objects under some illumination changes.

Four indexes have been generated for this dataset, considering according to case :

1. The 8 first order color invariants directly computed on the images ;
2. The 8 first order color invariants computed on images locally normalized, according to the technique proposed in [7] (IN1 normalization) ;
3. The 11 normalized second order invariants directly computed on the images (VN2 normalization) ;
4. The 11 normalized second order invariants computed on the images locally normalized (IVN2 normalization).

The precision/recall diagrams obtained for these indexes are presented in figure 8. We observe that the precision decreases rapidly when the basic invariants are applied directly on images. All the normalization approaches improve the retrieval process notably, but the best results are obtained for the VN2 normalization.

Whereas the contribution of second order invariants did not appear obvious at previous sections 3.1 and 3.2, these last results show that second order should be considered in order to get a point characterization robust to illumination variations. The image local normalization IN1 clearly reduces the illumination differences, but it seems to introduce some noise which cannot be eliminated even if the invariants normalization is added (IVN2).

3.4 Exploiting geometrical constraints or not ?

If n_q is the number of points of the image query and n the number of points in the database index, the algorithm of nearest neighbor search has a complexity in $O(n_q \times n)$. When the neighborhood constraints presented at section 2.1.4 are employed, this complexity increases notably. Let n_n be the average number of neighbors per 2D extracted point. Considering for example the angular constraint on gradients, which has got lower computation cost comparing to other appoaches, that complexity becomes quadratic by being multiplied by n_n^2 .

[1] Images provided by courtesy of the Computational Vision Laboratory, Simon Fraser University, USA, http://www.cs.sfu.ca/~colour/data/.

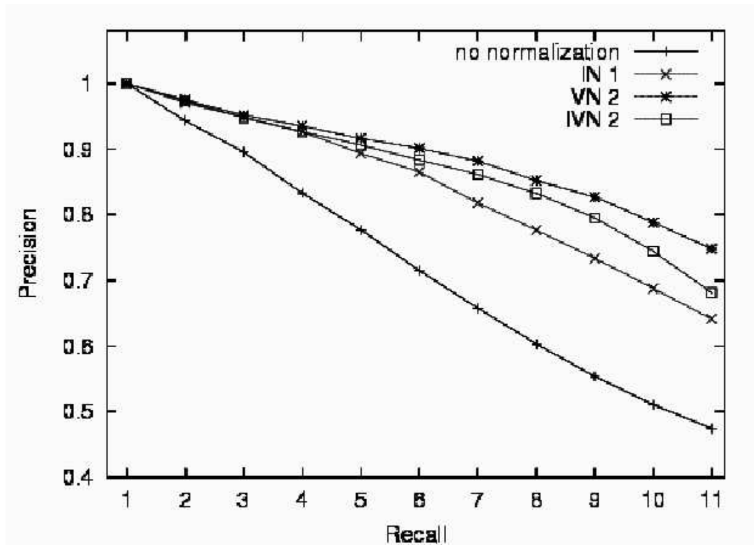


Figure 8: Precision/recall diagrams according to illumination changes normalization.

In this context, one can ask if this additional information is really essential for image indexing and retrieval applications. To get an answer, we re-indexed the benchmark database by considering first order invariants and adding neighborhood constraints. The constraints used were the basic constraint on the number of points matched in the neighborhood of the point plus its spatial relations by using the angular constraint on gradients. We do not present in this paper the precision/recall diagram obtained because it is roughly similar to the one obtained with photometric constraints only. Our explanation for this disappointing result is first that the benchmark database is not well suited for emphasizing the relevance of the neighborhood contraints, insofar as the sets of points globally have identical spatial distribution in each image, and second that the number of points per image (30) does not allow to build rich enough point neighborhoods. However we can conclude that the photometric information is clearly more relevant. Geometric constraints may allow to raise ambiguities in particular configurations of points, for example for insulated points or points gathered in a part of the image. To confirm this assumption, an experience was done on the particular case of object retrieval and is described below:

We considered images of 50 synthetic 3D objects, without background and which show the objects under many viewpoints[2]; some of them are presented at figure 9.

The set used for our experiments belong to a camera rotation around a vertical axis passing through the object and with 7 angles up to 180°. We could not use the dataset

[2]Images provided by courtesy of Michael J. Tarr (Brown University, Providence, RI, USA).

Figure 9: Subset of the 3D objects used for the automatic evaluation.

coil-100 of the Columbia University, as for previous experiments, because it was not possible to insulate the objects from background properly.

The image of each of the objects used which corresponds to a fixed point of view was superposed on the ones of the generic database used at previous section. No attention was paid to the images size, so that some objects may have been inserted partially. We obtained a database of 500 images with 50 of them containing the 3D objects seen under a specific viewpoint. The images related to the other viewpoints were not included in the dataset and were indexed separately.

The indexing process produced about 183000 points of interest characterized using first order color differential invariants plus neighborhood constraints. The automatic evaluation step then consisted in searching in this database the images containing the 3D objects, by considering as queries all the viewpoints available for these objects.

In figure 10, we present the ratio of first positive retrieval results for the 50 objects seen under 7 different viewpoints, that represents 350 queries at all. The retrieval was done first only by exploiting the photometric information contained in the color invariants and second by adding neighborhood constraints based on the angle between gradients.

The graphs obtained are function of the rotation angle traducing the relative position between the object superimposed in the database and the searched object. The zero angle traduces a difference between the background of the compared objects only. We notice the positive impact of the use of the neighborhood constraints on the retrieval. These ones allow to take into account the global spatial distribution of the query points in the retrieved images, in addition with the photometric similarity information. It is clear that the points of the sub-image query must be found relatively near to each other in the browsed images, what the only photometric constraints do not guarantee at all. Let us note finally that this improvement is relatively weak according to the complexity implied by this class of constraints (see considerations of section 3.4).

4 Conclusion

The purpose of this paper was to study the behavior of the different implementations of the color points of interest as regards practical cases like changes of viewpoint, image coding noise and illumination changes. We tried to study the relevance of the various parameters according to these configurations. Experiments on benchmark databases leaded to several observations. First, the very high sensibility to noise of the third order differential invariants was confirmed at sections 3.1 and 3.2 by considering viewpoint changes and the JPEG coding

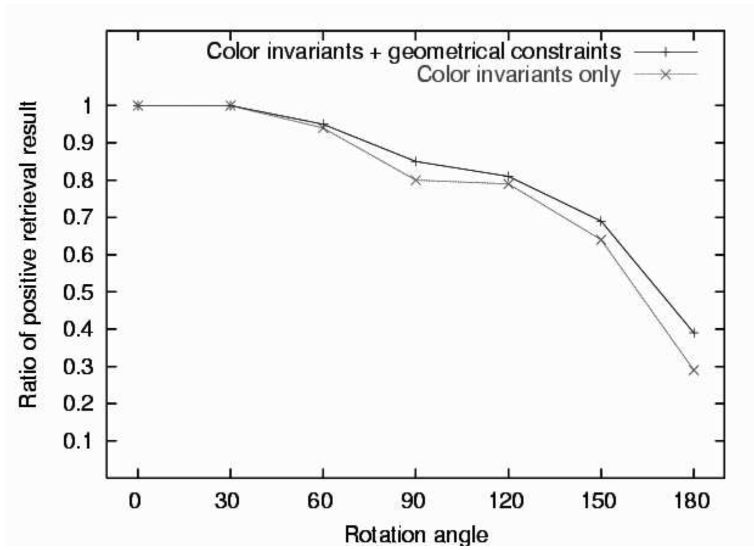
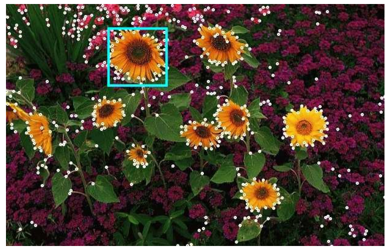


Figure 10: Results of object-based retrieval with/without neighborhood constraints.

format. The use of this very precise set of invariants proved relevant only in particular conditions involving images of high quality and differing from small viewpoints. It was shown too that second order magnitudes slightly produce better results than the basic description involving only first order ones. But the second interesting observation is that they appear more adequate than the others when considering illumination changes. Indeed in section 3.3, a comparison between image and invariants normalizations was done. This showed that mixing the invariants allow to notably reduce the effects of illumination changes on image retrieval. Consequently, this conclusion imposes the choice of invariants up to second order, seeing that first order quantities may not be used normalized. The third observation was about the neighborhood constraints, which are usually useful for stereoscopic applications. We saw here that the richness of the photometric information contained in the color invariants makes them not really relevant for global image indexing. However, this class of constraints got a positive impact for object retrieval (see section 3.4). One should take their relative complexity into account for implementing an efficient indexing system.

According to these considerations, we are developing an image indexing and retrieval platform based on color point of interest description, involving the normalized invariants up to second order and spatial relations based on the number of points matched in the point neighborhood. Figure 11 presents a typical example of sub-image retrieval using such local descriptors.

In parallel, our research works deal with the optimization of storage and retrieval costs of our feature vectors by considering multidimensional index structures, with the aim of fitting to scale. The main issue of this paper - the relevance of a particular class of differential invariants, is crucial here, since it is well-known that the choice of the index structures and of the search algorithms to consider largely depends on the size of the feature space.



Partial query on a rectangular area defined by the user.

Result of the query. The returned images are presented by decreasing order of similarity.

Figure 11: illustration of retrieval on sub-image using color points of interest.

References

[1] Adam Baumberg. Reliable feature matching across widely separated views. In IEEE *Conference on Computer Vision and Pattern Recognition*, pages 774–781, 2000.

[2] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color- and texture-based image segmentation using em and its application to content-based image retrieval. *Proc. Int. Conf. on Computer Vision (ICCV'98)*, 1998.

[3] T.O. Binford and T.S. Lewitt. Quasi-invariants : theory and exploitation. In *Proceedings of DARPA Image Understanding Workshop*, pages 819–829, 1993.

[4] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, 33:116–125, 1986.

[5] J. Fauqueur and N. Boujemaa. Image retrieval by regions: Coarse segmentation and fine color description. In *International Conference on Visual Information System (VISUAL'2002)*, Hsin-Chu, Taiwan, March 2002.

[6] V. Gouet and N. Boujemaa. Object-based queries using color points of interest. In *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL 2001)*, pages 30–36, Kauai, Hawaii, USA, 2001.

[7] V. Gouet and P. Montesinos. Normalisation des images en couleur face aux changements d'illumination. In *13ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle*, volume II, pages 415–424, Angers, France, 2002.

[8] V. Gouet, P. Montesinos, R. Deriche, and D. Pelé. Evaluation de détecteurs de points d'intérêt pour la couleur. In *Reconnaissance des Formes et Intelligence Artificielle*, volume II, pages 257–266, Paris, France, 2000.

[9] V. Gouet, P. Montesinos, and D. Pelé. A fast matching method for color uncalibrated images using differential invariants. In *Proceedings of the British Machine Vision Conference*, volume I, pages 367–376, Southampton, UK, September 1998.

[10] P. Gros. Color illumination models for image matching and indexing. In *Proceedings of* 15^{th} *International Conference on Pattern Recognition*, pages 580–583, Barcelona, Spain, September 2000.

[11] D. Hilbert. *Theory of Algebraic Invariants.* Cambridge Mathematica Library, Cambridge University Press, 1890.

[12] Jean-Michel Jolion and Stéphane Bres. Influence du codage jpeg sur des descripteurs d'images. *Traitement du signal*, 15(4):309–320, 1999.

[13] T. Lindeberg and J. Garding. Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure. *Image and Vision Computing*, 15(6):415–434, June 1997.

[14] J.M. Marie-Julie and H. Essafi. Using ifs and moments to build a quasi invariant image index. In *European Conference on Computer Vision*, 1998.

[15] Krystian Mikolajczyk and Cordelia Schmid. Indexing based on scale invariant interest points. In *International Conference on Computer Vision*, Vancouver, Canada, July 2001.

[16] P. Montesinos, V. Gouet, and R. Deriche. Differential Invariants for Color Images. In *Proceedings of* 14^{th} *International Conference on Pattern Recognition*, Brisbane, Australia, 1998.

[17] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–534, May 1997.

[18] J.R. Smith and S.F. Chang. Visualseek: A fully automated content-based image query system. In *ACM Multimedia*, pages 87–98, 1996.

[19] Zhengyou Zhang, Rachid Deriche, Olivier Faugeras, and Quang-Tuan Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78:87–119, December 1995.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399