

Content Based Image Retrieval Using Gradient Color Fields

F. Le Bourgeois

Equipe de Reconnaissance de Formes et Vision

I.N.S.A. de LYON - Bât 403

20 Avenue A. Einstein 69621 Villeurbanne Cedex FRANCE

Tél : (+33) 4 72 43 80 93 Fax : (+33) 4 72 43 80 97

E-mail : flebourg@rfv.insa-lyon.fr

Abstract

This paper presents a Content Based System Retrieval that uses gradient color fields as features. These features take into account both contour curvature and colors found in adjacent regions. This approach is dual from color region based methods. It allows a simple description of images using a coarse understandable sketch, which preserves rich information with a small amount of storage. This representation allows different type of queries such as : statistical or structural, global or partial and contour based. These features have been tested, using different modes on a very large image database from broadcast television. Results obtained from our system is presented and discussed.

1. Introduction

The increasing number of image that are not indexed, in places such as the World Wide Web or video databases, oblige us to improve capabilities of Content Based Images Retrieval systems (CBIR). In particular, we have to find the best features to describe images with a reduced amount of storage containing a maximum quantity of relevant information. Two solutions have been studied in previous work :

- **Description in a high semantic level:** This solution describes directly the content of images textually, after an automatic recognition. This task is difficult to achieve without any previous knowledge about images contents. In some cases, we can recognize directly some contents like faces, cars, captions and particular objects which have a model. It's also possible to classify images into opposite categories (Indoor/outdoor, day/night, Synthesis/natural images). The limitation of such approach make it difficult to index unconstrained natural images .
- **Description in a low semantic level:** This approach describes images by numerical features that are compared using an appropriate similarity measure. This approach is currently used by several CBIR systems like Qbic[1], Virage [3], Candid [4] or Photobook [5]. Queries are made by example using an image or a

sketch. Users generally find it difficult to understand some features like colors, textures, and contour shapes. Moreover some adjustment of abstract parameters for the comparison is also another difficult problem to figure out.

The choice of features is fundamental for any CBIR system performance. These features can be classified into three categories: colors, texture and shape. Another category concerns methods which cannot be classified into previous categories like eigenimages, coarse contour of images consider as a primal sketch [5], features based on interest points [6].

The work on primal sketch published by Kato [5] constitutes the origin of our work. The author shows that for a CBIR system, a precise contour is not necessary. A simple rough sketch based on a coarse contour is sufficient for a query by example. The comparison is made by matching rough map of contour with a sketch provided by the user. This approach is similar to D. Marr [7] works on primal sketch and human vision. But this work can be criticized in many points like the lack of information about colors, the use of only high gradient part of images as unique feature. We present in this paper an improvement over Kato's work. We believe that a primal sketch of images is a natural and efficient description of unknown images easily understandable by any users who can modify it during a feedback process. Compare to Kato's work, we introduce more information about local curvature of contours by using gradient direction and colors provided by the color gradient vector. In the first part, we present our features extraction and our approximation of color gradient by using another matrix normalization. In the second part we describe different types of feature comparison algorithm and finally we give some results on an large database from a broadcast television videos.

2. Features extraction

A frontier-based description is almost equivalent to a region-based representation. The query "*Find images which contain a red disk on a yellow background*" can be

reformulated in another way as "find images containing a contour of a disk having gradient oriented to the center and changing colors from yellow to red". In this example, the orientation of the gradient brings important information about the shape of the object (gradient vectors along a circle are always oriented to the center) and some work already used it for contour matching [8]. We obtain a more realistic description of image contents using contour detection than a region-based segmentation. We do not need a perfect closed contour for a query by contents and the implementation of a frontier-based representation is easier to achieve.

We propose to keep the first partial derivatives from both sides of the contour in order to preserve color information and contour curvature. For this task, we need to compute the gradient for color images. The color gradient estimation is a complex mathematical problem, which have been studied in some work [9][10][11][12][13][15]. Let a color image be a mapping f defined by :

$$f : IN^2 \rightarrow IN^3 \\ (x, y) \mapsto (R(x, y), V(x, y), B(x, y)) \quad \Delta f = \begin{pmatrix} \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} \\ \frac{\partial V}{\partial x} & \frac{\partial V}{\partial y} \\ \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{pmatrix}$$

Color Gradient Δf can be represented by the Jacobian matrix J of 3x2 elements. The problem consists to approximate the direction $\theta(\Delta f)$ and the normalization $\|\Delta f\|$ of this matrix. Differential geometry framework and tensorial calculation is generally used. We can recognize in this physician approach the mathematical concept of Euclidean matrix normalization also named the spectral radius of the Jacobian matrix J . The gradient normalization is given by the most important eigenvalue and the direction is provided by eigenvectors in matrix P which describe a geometrical rotation. It is obtained with a eigen decomposition of the 2x2 matrix calculated from the inner product of the Jacobian matrix $J^t J$ (1).

$$J^t J = \begin{pmatrix} \frac{\partial R}{\partial x} & \frac{\partial V}{\partial x} & \frac{\partial B}{\partial x} \\ \frac{\partial R}{\partial y} & \frac{\partial V}{\partial y} & \frac{\partial B}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} \\ \frac{\partial V}{\partial x} & \frac{\partial V}{\partial y} \\ \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial R^2}{\partial x} + \frac{\partial V^2}{\partial x} + \frac{\partial B^2}{\partial x} & \frac{\partial R}{\partial x} \frac{\partial R}{\partial y} + \frac{\partial V}{\partial x} \frac{\partial V}{\partial y} + \frac{\partial B}{\partial x} \frac{\partial B}{\partial y} \\ \frac{\partial R}{\partial y} \frac{\partial R}{\partial x} + \frac{\partial V}{\partial y} \frac{\partial V}{\partial x} + \frac{\partial B}{\partial y} \frac{\partial B}{\partial x} & \frac{\partial R^2}{\partial y} + \frac{\partial V^2}{\partial y} + \frac{\partial B^2}{\partial y} \end{pmatrix}$$

$$\|\Delta f\| = \max\{\lambda_1, \lambda_2\} \quad J^t J = P \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} P \quad (1)$$

We introduce a different approximation from an another matrix normalization (2). We define the differential dx, dy from the maximum absolute value of partial derivatives, for each channel (3), but the sign of the derivatives is preserved in dx, dy (see $Cmax$ fct. (4)).

$$\|\Delta f\| = \left\| \begin{pmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \\ g_{3,1} & g_{3,2} \end{pmatrix} \right\| = \text{Max}_{j=1,2} \left\{ \text{Max}_{i=1,3} \{|g_{i,j}|\} \right\} \quad (2)$$

$$dy = C \max \left\{ \frac{\partial R}{\partial y}, \frac{\partial V}{\partial y}, \frac{\partial B}{\partial y} \right\} \quad dx = C \max \left\{ \frac{\partial R}{\partial x}, \frac{\partial V}{\partial x}, \frac{\partial B}{\partial x} \right\} \quad (3)$$

$$C \max \{g_k\} = \{g_k / |g_k| > |g_p| \forall k \neq p\} \quad (4)$$

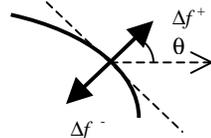
The direction is given by (5) and gradient normalization is provided by (2) which is equivalent to (6)

$$\theta = \arctg \left(\frac{dy}{dx} \right) \quad (5) \quad \|\Delta f\| = \text{Max}\{|dx|, |dy|\} \quad (6)$$

Our approximation of the color gradient orientation and normalization provide almost same results compare to previous expensive methods. It is important to notice that our proposal is different from taking the derivatives of the maximum of the three RGB channels because maximum of derivatives is not equal to the derivative of a maximum. Moreover it is easier to implement our approximation in a Canny optimal contour Detection [14]. We estimate the color derivatives dx, dy using (7)

$$dx = C \max \left\{ R \otimes \frac{\partial G_\sigma}{\partial x}, V \otimes \frac{\partial G_\sigma}{\partial x}, B \otimes \frac{\partial G_\sigma}{\partial x} \right\} \\ dy = C \max \left\{ R \otimes \frac{\partial G_\sigma}{\partial y}, V \otimes \frac{\partial G_\sigma}{\partial y}, B \otimes \frac{\partial G_\sigma}{\partial y} \right\} \quad (7)$$

From dx and dy , we can apply classical algorithm to trace edge pixels and suppress false edges. For each point of the contour, we define left gradient Δf^+ and right gradient Δf^- according to directional partial derivatives.

$$\nabla f^+ = \begin{pmatrix} dR^+ \\ dV^+ \\ dB^+ \end{pmatrix} \quad \nabla f^- = \begin{pmatrix} dR^- \\ dV^- \\ dB^- \end{pmatrix}$$


In practice we prefer to store the average color for each side of the contour than the derivatives, the absolute values are more explicit to users for a query. We keep six values in a feature vector $X = \{R^+, R^-, V^+, V^-, B^+, B^-\}$. For $C = \{R, G, B\}$ we calculate the features using with ρ proportional to the gradient normalization (8)

$$C^+ = \frac{1}{\rho} \sum_{h=1}^{h=\rho} C(x + h \times \cos(\theta), y + h \times \sin(\theta)) \quad (8)$$

$$C^- = \frac{1}{\rho} \sum_{h=1}^{h=\rho} C(x + h \times \cos(\theta + \pi), y + h \times \sin(\theta + \pi))$$

We illustrate the process on noisy compressed JPEG image from the French television. Figure 1 shows all color gradient vectors and the image reconstruction by filling regions from the gradient direction with colors stored in feature vector X . We notice that color gradient fields contain enough information to reconstruct a coarse image. We reduce the number of features in order to limit the storage by eliminating adjacent identical color gradients. We preserve gradient at some specific part of the contour corresponding to singular points like sharp edges, corners or high curvature changes. We also preserve gradients at regular intervals to describe linear shapes. From the example Figure 1 we reduce the number of gradient to 648 gradients for a 384x288 size of image from MPEG1



Figure 1: Original image, color gradient fields, image reconstruction and fields simplification

video. For all images from the same origin, we find an average of 500 gradients, with a minimum of 100 gradients for simple images to more than 1000 gradients for very complex images. For each gradient we store nine fields coded in one byte each:

- (x,y) relative position of the gradient in percentage
- $\theta(\Delta f)$ gradient orientation in degree
- $X = \{R^+, V^+, B^+, R^-, V^-, B^-\}$ Color information

The size of each image in our database is about 4Kbytes. The features are stored in a dictionary trie nodes referenced by a tri-dimensional index uses gradient location and orientation $\omega = (x, y, \theta)$. This structure allows to reach quickly any gradient from a given location and orientation independently of all images of the database. This structure is optimized for speed and storage. Moreover it is suited for every type of queries based on color gradient fields. Each color information X referenced by its location and orientation ω has two links. The first pointer links horizontally all other gradients having the same location and orientation for all images of the database. The second pointer links vertically gradients from the same image which share the same contour. This second link permits structural queries which use contour constraints.

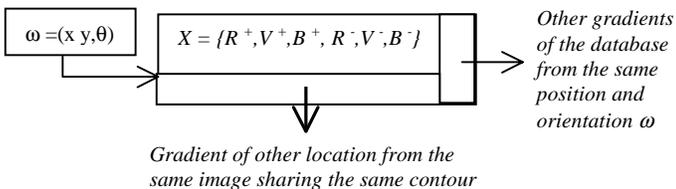


Figure 2: Data structure and index

Compared to other work based on interest points [6][12], our approach is almost different. Our feature selection use

spatial singularities and high curvature of the contour. In the opposite, interest point based on corner detection like Harris method [15] localize corner singularities in the multidimensional edge map. The finesse of our representation depends on several parameters. A set of parameters concerns the contour tracking (standard deviation of the Gaussian, low and high threshold). Another set of parameters is about the suppression of redundant information and the selection of gradient located on singularities points or high curvature of the contour. The adjustment of these parameters depends of the size and quality of images, the coarseness of the representation wished and the storage limitation.

3. Features matching

All different type of queries use the same matching process between two images I_1 and I_2 given by (9). We compute the statistical average distance between gradients which have the best match under the constraint that their localization ω are similar, in other words that ω_1 is in the neighborhood of ω_2 ($\omega_1 \in N(\omega_2)$) or reciprocally.

$$Score(I_1, I_2) = \frac{1}{M} \sum_{\substack{\omega_1 \in N(\omega_2) \\ \omega_2 \in N(\omega_1)}} \text{Min}(|X_1 - X_2| + b|\omega_1 - \omega_2|)$$

$$M = \text{Card}\{\omega_1 \in N(\omega_2) \cup \omega_2 \in N(\omega_1)\} \quad (9)$$

The user can set $b=0$ if he do not wants to take into account the difference of features localization and orientation in the global score. It makes it possible to find similar images independently of homogeneous transformations and scale variations. All query modes are defined by different neighborhoods:

- **Statistical global query**

This query compares statistically all gradients which have a direction and a localization similar in a range defined by the user: The ranges $\epsilon_x, \epsilon_y, \epsilon_\theta$ allow the user to define translation and rotation tolerance. These parameters influence directly the computation time.

$$\omega_1 \in N(\omega_2) \Leftrightarrow \{ |x_1 - x_2| \leq \epsilon_x \cap |y_1 - y_2| \leq \epsilon_y \cap |\theta_1 - \theta_2| \leq \epsilon_\theta \}$$

- **Statistical partial query**

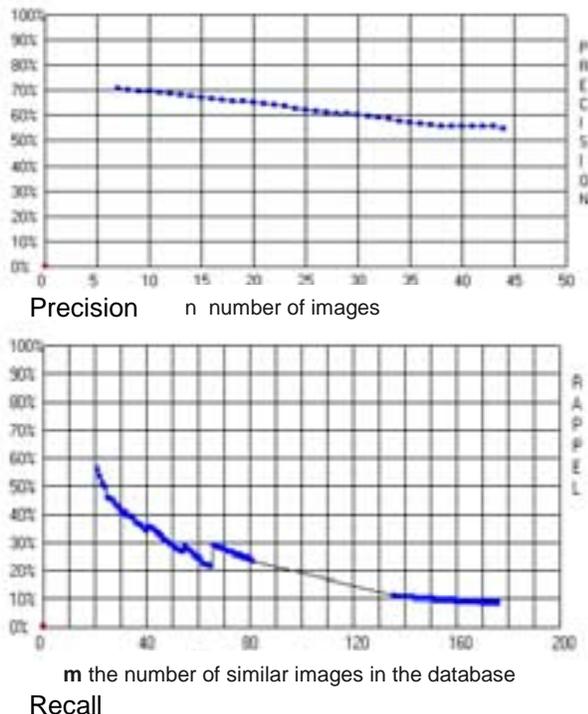
Color gradient fields makes partial queries possible. This type of query is very useful because user can define the region of interest and neglect parts which are not of interest to the match. We keep the same neighborhood from the previous mode but we limit the computation on gradients to the area of interest defined by user.

- **Query based on contour constraints**

For this mode, we restrict gradients to belong to the same contour. It is not a real structural comparison since we measure statistically the number of relevant matches of gradients having similar orientation and localization along a same contour.

4 Results

In this paper, we present the results only on the first type of queries. We use a private database of 27718 images from the national French television. These images were grabbed every 12 seconds during 24 hours from four different channels and stored in JPEG compressed at 75%. These images present a great variety of emissions (movies, news, advertisements, talk shows, and sports...). We use the conventional measures of Precision ($P=r/n$) and Recall ($R=r/m$) where r is the number of right images returned by the system, n the number of desired images and m the number of similar images effectively existing in the database. These results are obtained with optimal range $\epsilon_x, \epsilon_y, \epsilon_\theta$ of translation rotation and scale variation found exhaustively. The optimal range, for a P/R optimal response, corresponds to 25-33% of translation and scale variation and $\pm 22^\circ$ of rotation tolerance. The computation time varies according to the translation, scale and rotation tolerance. With optimal tolerance described previously, the system need less than a minute to obtain the result of a query. Results are satisfactory with highly structured images containing geometrical information (sports, news, indoor shows). Results decrease for images containing random information or having no precise sketch (uniform color image or images with no precise contour). On the other side, very good results have been measured ($P \approx 1$) with queries about textured images (grass, stone, wood) outdoors views and images of captions.



5. Conclusion

In combination with other features, color gradient fields make a good representation of images, which have interesting capabilities to describe coherent textures and geometrical information. Images are represented by a color coarse sketch, which can be easily interpreted by any user. Color gradients contain a great quantity of information about curvatures of contours and colors of adjacent regions. To improve the feature extraction process, we have introduced another multidimensional gradient normalization and orientation that is easier to compute. Results are encouraging, especially for real images from television that contain a lot of noise due to motions and Jpeg compression.

6. References

- [1] W. Niblack et al., *The Qbic Project: querying images by content using colour, texture and shape*. In SPIE Proc. Storage and retrieval for Image and Video databases, 1993, pp. 173-181.
- [3] A. Hampapur, and al., *Virage Video Engine. Storage and Retrieval for Image and Video Databases* (SPIE) 1997 pp. 188-198.
- [4] P. Kelly, M. Cannon, D. Hush, *Query by image example: The comparison algorithm for navigating digital Image databases (CANDID) approach*. SPIE proc Storage and retrieval for Image and video databases pp. 238-249, 1995
- [5] Toshikazu Kato, Kyoji Hirata, *Query by visual Example*, in EDBT'92, pp. 56-71, 1992
- [6] C. Schmid, R. Mohr, Local Grayvalue invariants for Image retrieval, IEEE PAMI, Vol 19, No 5, May 1997, pp. 530-532
- [7] D. Marr, *Vision, a Computational Investigation into the Human Representation and processing of visual Information* in W. H. Freeman and co, ISBN 0-7167-1567-8, 397 p.
- [8] A. Del Bimbo, P. Pala, *Visual Image Retrieval by Elastic Matching of User Sketches*, IEEE PAMI, Vol 19, No 2, Fevr 1997, pp. 121-132.
- [9] M. Chapron, *A new Chromatic Edge Detector used for Image Color Segmentation*, 11th IAPR, The Hague, Vol 3, 1992, pp. 311-314.
- [10] R. Nevatia, *A color Edge Detector and its use in Scene segmentation*, IEEE trans. On System Man. and Cybernetics, vol 7, 1997, pp. 820-826.
- [11] G. Sapiro, D. L. Ringach, *Anisotropic Diffusion of multi-valued Images with application to color filtering*, IEEE PAMI (5)11, 1996, pp. 1582-1586.
- [12] P. Montesinos, V. Gouet, R. Deriche, *Differential invariants for color Images*, 14th IAPR, Brisbane, 1998, Vol 1, pp. 838-840.
- [13] S. di Zenzo, *Gradient of multi-Images*, CVGIP, 33 pp. 116-125, 1986.
- [14] J.F. Canny, *A Computational Approach to Edge Detection*, IEEE PAMI, vol 8, n°6, pp. 679-698, 1986
- [15] C. G. Harris, M. Stephens. *A Combined corner and edge detector*, Proc 4th Alvey Vision Conference, pp. 147-151, 1988.