High-Level Clothes Description Based on Colour-Texture and Structural Features

Agnés Borràs, Francesc Tous, Josep Lladós, and Maria Vanrell*

Computer Vision Center - Dept. Informàtica
UAB Bellaterra 08193, Spain
{agnesba,ftous,josep,maria}@cvc.uab.es
http://www.cvc.uab.es

Abstract. This work is a part of a surveillance system where content-based image retrieval is done in terms of people appearance. Given an image of a person, our work provides an automatic description of his clothing according to the colour, texture and structural composition of its garments. We present a two-stage process composed by image segmentation and a region-based interpretation. We segment an image by modelling it due to an attributed graph and applying a hybrid method that follows a split-and-merge strategy. We propose the interpretation of five cloth combinations that are modelled in a graph structure in terms of region features. The interpretation is viewed as a graph matching with an associated cost between the segmentation and the cloth models. Finally, we have tested the process with a ground-truth of one hundred images.

1 Introduction

In many application fields large volume of data appear in image form. The Content-Based Image Retrieval (CBIR) is the Computer Vision area in charge to handle and organize this great volume of data due to its visual content. Image retrieval from databases is usually formalized in terms of descriptors that combine salient visual features such as colour, texture, shape and structure. For any given feature there also exists multiple representations that characterize it from different perspectives. The reviews of Huang [11] and Forsyth [7] expose a wide variety of feature representations and image retrieval strategies.

This work is focused on the development of a content-based retrieval system where the image classification is done according to the presence and description of a certain object. The process involves two steps: an image segmentation and a region based interpretation. In the first step, the information of the segmented image is organized as an attributed graph which features characterize the regions and their relationships. We define certain operators that, following a split-and-merge scheme, allow the graph to evolve until finding the final solution. Image

* This work has been partially supported by the project TIC2000-0382 and the grant 2002FI-00724.

F.J. Perales et al. (Eds.): IbPRIA 2003, LNCS 2652, pp. 108–116, 2003.

segmentation techniques can be roughly classified into four groups: pixel based, boundary based, region based and hybrid techniques. Some understanding surveys on image segmentation are those of Haralick and Shapiro[9] and Muñoz[13]. Our segmentation strategy is classified as a hybrid method for combining clustering in the colour space, colour homogeneity and edge detection. In the second step of our process, image interpretation, the structure of the segmented regions is matched against a set of models of objects. These models are also represented as graphs that contain features such as colour, texture, size, shape and position. Hence, the interpretation step is performed as a matching procedure between the graph of the segmented image and the graph of the model objects. The best matching solution is chosen due to a cost measure provided by the matching operations on the model features.

We have tested our system by integrating it as a retrieval module of a general surveillance application. This application performs image retrieval in terms of people appearance and acts as a control mechanism of the people that enters in a building. It automatically constructs an appearance feature vector from an image acquired while people is checking-in in front of an entrance desk. This way, the system analyses some person characteristics, such as the height, the presence of glasses or the clothing, and stores the result in a database. Thus, a graphic based interface allows the security personnel of the building to perform an image retrieval of the registered people by formulating queries related on their appearance. The objective of our work is centred in the module that provides an automatic description of the people clothing. This description is given in natural language in terms of colour, texture and structural composition of the garments.

In the literature we can find several examples of strategies that, like the one which we have developed, combine region features and graph structure for database indexing [6][14]. However, in the concrete aim of the clothing description, the most similar approach consists in the Changs development of a computer-aided fashion design system [3]. However, this approach treats the clothing segmentation process but does not treat the interpretation one.

The paper is organized as follows: in section 2 we detail the image segmentation according to its graph-modelling and its strategy. In the section 3 we present how we model the clothing compositions as another graph of features and how we perform the matching to interpret the clothing regions. Next, in the section 4, we expose an example of the retrieval behaviour of our module. Finally, in the sections 5 and 6 we present some results and conclusions respectively.

2 Image Segmentation

2.1 Segmentation Modelling

Graph Representation. We model an image I as a set of non-overlapping regions R structured by an attributed graph G. The graph G is formed by a set of nodes N, a set of edges E, and two labelling functions over these nodes and edges. While each node identifies an image region r, each edge represents a relation between two regions r_i, r_j. The graph is also provided with two labelling functions, L_N and L_E. They are in charge to obtain and store the feature information F_N and F_E that identifies the nodes and edges respectively.

$$G = (N, E, L_N : N \to F_N, L_E : E \to F_E)$$

Node Features $F_N = \{BB(n), A(n), H(n), E(n), AC(n), AI(n), T(n)\}$: A region is described with its bounding box (BB), the area (A), the colour histogram (H), the edge presence (E), the average chromaticity (AC), the average intensity (AI), and the texture presence (T).

Edge Features $F_E = \{D(n_i, n_j), NH(n_i, n_j)\}$: The region relations are defined by the neighbourhood information (NH) and a similarity distance (D). In the next section 2.2 we detail how D is computed from the node features.

Graph Edition Operations. We define two graph operators that work over the graph structure and allow it to grow and to diminish. These operators are the fusion operator γ_F and the division operator γ_D. After a step of graph expansion or contraction, they are in charge to recalculate F_N and F_E and restructure G (remove obsolete edges, etc).

2.2 Segmentation Process

Algorithm Steps. As we illustrate in the Figure 1, our segmentation algorithm is a process that consists in three steps: initialisation, split and merge.

Starting from the source image I and a mask of the zone we want to segment, we create the initial graph G as a unique node. Then we expand G in two phases corresponding to a discrimination of the textured areas and a breaking of the plain ones. Thus, the division operator $\gamma_D(G)$ acts over the graph nodes due to some predefined split criteria SC based on the node features F_N. Finally we

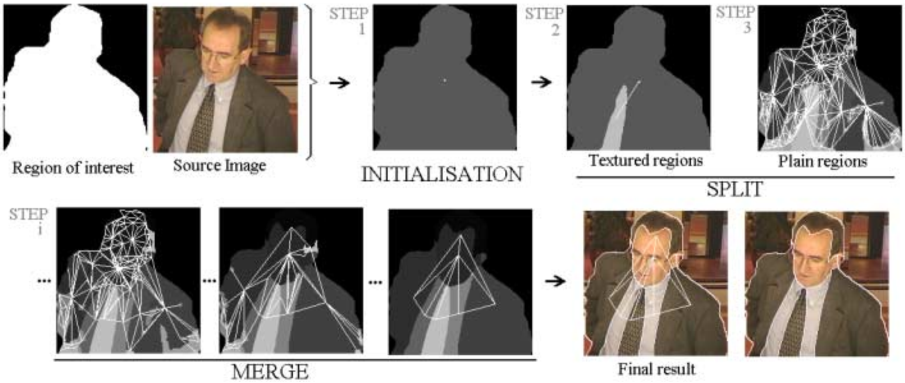

Fig. 1. Segmentation process guided by a graph structure

apply iteratively the fusion operator $\gamma_F(G)$ due to some merge criteria MC that deal with the edge features F_E of the graph. Next we expose the criteria we follow to apply the operators in the split and merge steps.

Split Criteria. We deal with homogeneity measures on the node features.

Texture Discrimination. We discriminate the texture zones by applying a statistical strategy inspired in the work of Karu[12] and in the MPEG-7 texture descriptor[5]. The general idea of our process is to consider as textured regions those image zones with a certain amount of area that present a high density of contours checked at certain frequencies. The exact detection steps are graphically showed in the Figure 2. The node feature E stores the edge information, and T indicates the texture presence.

Plain Regions Split. We apply a pixel-based technique that consists in a clustering of the colour space. A plain region will be formed by all the connected pixels in the image that belong to the same colour cluster. We have used the octree quantization algorithm of Gervautz and Purgathofer[8] that, given a number of colours nc, provides the palette of the ncth most usual colours of the image. This adaptability is very interesting to avoid the under segmentation when we deal with garment combinations of very similar colours. The quantization information is stored in the node feature H.

Merge Criteria. We allow the fusion of two adjacent regions if their similarity feature D is under a certain threshold. Being this value a measure between 0 and 1, the fusion operator will be applied iteratively to the pair of neighbouring regions with minimum distance.

Plain Regions Similarity. The shadows provided by the clothes folds are viewed as intensity changes that become especially critical in the case of the plain regions. Thus, we have developed a similarity distance that gives more tolerance to the intensity variations and allows the presence of progressive and smooth intensity degradation in a region. The similarity measure is computed by a combination of a chromatic distance and an intensity distance. The chromatic distance is computed from the AC node features as the Euclidean distance between the colour means on the chromatic plane. When two regions are adjacent, the intensity distance ID is computed from the E node features as the rate of edge

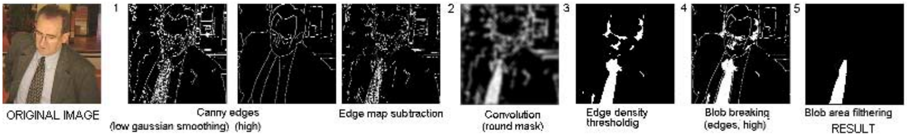

Fig. 2. The five steps of the texture discrimination process

pixels in the common boundary. Otherwise, we calculate ID as the Euclidean distance between the average intensity AI of the regions.

Textured Regions Similarity. We use the histograms of the two regions (H) as their texture descriptors. We use a similarity metric that treats simultaneously the distances of the histogram rates and the distances of the colours that they represent. This measure is commonly used for region based image retrieval and is defined as a similarity colour descriptor in the MPEG-7[5] encoding.

3 Interpretation

3.1 Interpretation Modelling

We attempt to distinguish between five types of clothing compositions that are combinations of two garments (buttoned or unbuttoned) and a tie. We understand the garments of a class composition as ordered layers from the most external to the most internal. For example we describe a person wearing an unbuttoned black jacket and a blue shirt, like a structure of two layers, the first black and the second blue. In terms of garment regions this can be seen as two black outer regions and one blue inner region.

We describe a clothing composition by a an ideal model structured as an attributed graph G_M where the nodes N_M represent the garment regions gr and the edges E_M their relationship (see Figure 3).

$$G_M = (N_M, E_M, L_{N_M} : N_M \to F_{N_M}, L_{E_M} : E_M \to F_{E_M})$$

Model Node Features $F_{N_M} = \{A(n_m), S(n_m), CL(n_m), CH(n_m)\}$: The model regions are defined by its ideal area (A) understood as the area rate with respect to the whole object. The region limits are analysed in order to identify a certain shape (S). Furthermore, we can set some colour restrictions by forcing the region to have a certain colour homogeneity (CH) and being this colour homogeneity of a certain label (CL) such as skin, grey, blue, pink, etc. We use the 25 colour label classification proposed by the colour naming method of Benavente[1].

Model Edge Features $F_{E_M} = \{SP(n_{mj}, n_{mk}), SI(n_{mj}, n_{mk})\}$: We need to add some similarity restrictions (SI) to those regions that, even thought of being apart, belong to the same garment (for instance the two regions that describe an unbuttoned jacket). We indicate the relative spatial positions between two

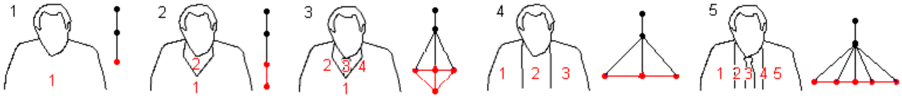

Fig. 3. Modeling of the five possible clothing compositions

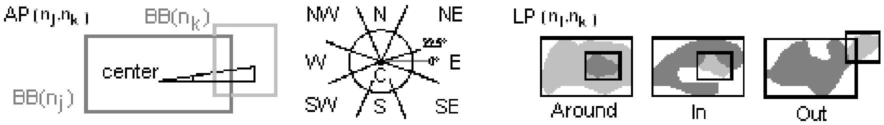

Fig. 4. Spatial position labels: [AP∈{N,NW,W,SW,S,SE,E,NE,C}, LP∈{I,A,O}]

regions (SP) with a combination of two labels $[AP, LP]$. These labels are obtained from the region bounding boxes and are inspired in the iconic indexing techniques of Rs-String [10] and 2D String [4]. Figure 4 show them graphically.

3.2 Interpretation Process

The interpretation process consists in evaluating all the possible mapping solutions between a segmentation graph G and each model graphs G_M. Minimizing a cost value associated to matching operations chooses the best result. The interpretation process applies an n-to-one mapping between the image regions and the model regions. It also allows an image region not to take part in the solution. The procedure pretends to avoid the over segmentation problem and reject those intrusive regions (bags, wallets, etc.) that do not belong to the clothing.

Matching Cost. We compute the mapping between a graph G and a model G_{Mi} due to some cost functions. These functions evaluate how the node features and the edge features of the model are preserved when they are mapped to the image ones. The functions δ_A, δ_{CH}, δ_{CL}, and δ_S, evaluate $F_{N_{Mi}}$, and the functions δ_{SI} and δ_{SP}, evaluate $F_{E_{Mi}}$. Let us name $\delta_{F_{N_M}}(\{n\}_i, n_{mi})$ and $\delta_{F_{ME}}(e_i, e_{mi})$ the combination of the node costs and edge costs respectively. In a higher level, the function δ joins and weights them with the parameters, $\alpha_{N,i}$ and $\alpha_{E,i}$. These parameters enhance the significance of a model part or of a relationship.

$$\delta(G_k, G_{Mi}) = \sum_{i=1}^{\#N_{Mi}} \alpha_{N,i} * \delta_{F_{N_{Mi}}}(\{n\}_i, n_{mi}) + \sum_{i=1}^{\#E_M} \alpha_{E,i} * \delta_E(e_i, e_{mi}) \qquad (1)$$

Next we define in a general way how we calculate the costs related with each feature. For more details, see Borràs[2]. The functions δ_A, δ_{SI} and δ_{CH} provide cost measures that vary in a range of goodness from 0 to 1 in reference to the area (A), similarity (SI) and colour homogeneity (CH) features. The area cost is computed as the ratio of the difference between the $\{n\}_l$ and mn_l areas. The similarity and cohesion costs are computed as the mean of the colour-texture distances defined in the section 2.2. In relation to the features with boolean

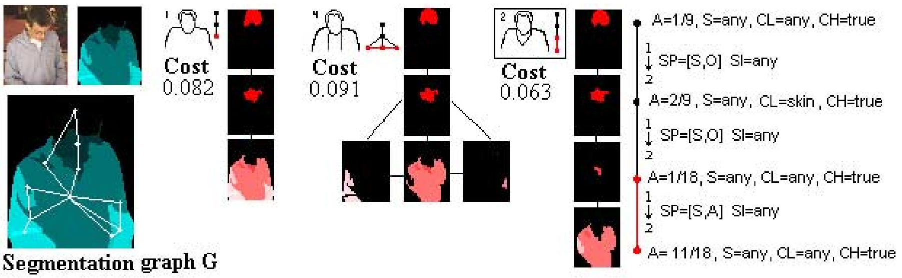
1
Cost
0.082
4
Cost
0.091
2
Cost
0.063
A=1/9, S=any, CL=any, CH=true
1 ↓ 2 SP=[S,O] SI=any
A=2/9, S=any, CL=skin, CH=true
1 ↓ 2 SP=[S,O] SI=any
A=1/18, S=any, CL=any, CH=true
1 ↓ 2 SP=[S,A] SI=any
A= 11/18, S=any, CL=any, CH=true
Segmentation graph G

Fig. 5. Given a segmentation graph G the figure shows the three best matching for the graph-modelled classes: 1,4 and 2. There is no result for 3 and 5 due to the absence of tie shape. The image is classified as Class2 since it has the lowest matching cost

properties, their costs are set to 0 or ∞ according to its accomplishment. The function δ_{SP} checks the space labelling (SP) and δ_{CL} examines the colour labelling of a region (CL) using the colour naming method [1]. Finally δ_S analyses the shape with synthetic tie mask.

Matching Process. From graph G and a model graph G_{Mi}, we make an expansion in a depth-search priority of a decision tree. Each tree level represents the mapping of a region model with a set of segmented neighbouring regions. Each tree node has associated a cost mapping of the partial solution. At each step, we only expand the nodes with a cost value ≤ 1. When the process is done for all G_M we choose the segmentation solution G_i^k with minimum cost $C_{i,k}$ ≤ 1. Applying the matching process to the whole models and observing the minimum value of each best mapping, we decide the class classification of the clothing composition. Figure 5 exemplifies a graph matching solution.

4 Example

We exemplify the behaviour of our method in front of a query formulated against a database. This database contains the clothing descriptions that our method has generated from a set of 100 test images, as well as, the colour labelling of identified garments [1]. Then, we try it out with two queries which results are showed in the Figure 6. A first query would be formulated as: ”We search a person wearing a clothing composition of two layers: the first opened, the second closed; with indistinct colour for the first layer, and white for the second layer”. Then, a second query could refine the previous one adding a colour restriction for the first layer as: ”...with black colour for the first layer...”.

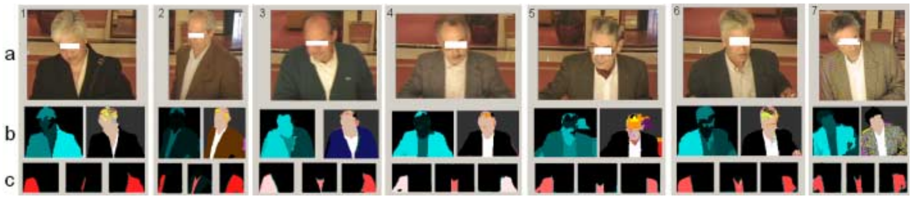

Fig. 6. Image retrieval: 1 to 7. Refined retrieval: 1,4,5 and 6. (a) Original image (b) Segmentation and colour naming (c) clothing regions of the structure identification

5 Evaluation of the Results

Starting from a set of one hundred images $\{I^j\}_{\{j:1..100\}}$ taken from a real environment, we have evaluated the whole process and their intermediate steps. We have chosen an empirical discrepancy method based on a set of ground truth information. We have used a synthetic segmentation of the images $SG = \{G^j\}$ and a manual labelling of their structure $SG_M = \{G^j_{Mi}\}$. According to them we have extract some statistics over two sets of structure results that we have obtained form two experiments. The first set, RG^I_M, is obtained by running our method starting from the original images. The second set, RG^{SG}_M, is obtained by running it from the synthetic segmented images.

Global Evaluation. Running our method form the original images we have obtained a success of 64% on the clothing classification ($SG_M \cap RG^I_M$=64%)

Segmentation Evaluation. We have compared the success on the structure identification starting from the original images and starting from the synthetic ones. Then we have obtained that $SG_M \cap RG^I_M$=64% and $SG_M \cap RG^{SG}_M$=69%. Therefore we observe that the automatic segmentation influences the process by incrementing the structure misclassification in a rate of 5%. This way, we can evaluate the segmentation success in a rate of 92.75%.

Structure Description Evaluation. As we have seen in the previous results, the structure description method can be evaluated with a 69% of success without the segmentation influence. The mean reasons that introduce this 31% of error are given by altered positions if the person in the image scene and severe occlusions on the cloth zones provided by external objects.

6 Conclusions

We have developed a content-based image retrieval strategy that we have applied to a problem of people clothing identification. Our process consists in two stages, image segmentation and interpretation, both guided by a graph structure. Even thought the difficulties that the clothes segmentation carries (the shadows of their folds, the irregular textures, etc.), our segmentation method fulfils satisfactorily the objective. To perform the interpretation step, we have modelled five types of clothing compositions according to some region features. We use several cost functions to evaluate the best matching between the regions of the segmented image and the ideal regions of the clothes composition models. The process attempts to overcome the over segmentation problem by allowing an n-to-one region mapping. Our strategy can be adapted to recognize and describe in terms of regions any object due to their colour, texture and structure features.

References

[1] Benavente, R., Olivé, M. C., Vanrell, M., Baldrich, R.: Colour Perception: A Simple Method for Colour Naming. 2n Congrés Català d'IA, Barcelona (Spain) (October 1999) 340-347 112, 114

[2] Borràs A.: High-Level Clothes Description Based on Color-Texture Features. Master Thesis. Computer Vision Center - Dept. Informàtica UAB (September 2002) 113

[3] Chang, C. C., Wang, L. L.: Color Texture Segmentation for Clothing in a Computer-Aided Fashion Design System. IVC Volume 14, Number 9 (1996) 685-702 109

[4] Chang, S. K., Shi, Q. Y., Yan, C. W.: Iconic Indexing by 2-D Strings. IEEE Trans. on PAMI Volume 9, (May 1987) 413-428 113

[5] Choi, Y., Won, C. S., Ro, Y.M, Manjunath, B. S.: Introduction to MPEG-7: Texture Descriptors (2002) 213-230 111, 112

[6] Chen Y., Wang J.: A region-based fuzzy feature matching approach to content-based image retrieval. IEEE Trans. on PAMI Volume 24 (2002) 109

[7] Forsyth, D. A., Malik, J., Fleck, M. M., Greenspan, H., Leung, T. K., Belongie, S., Carson, C., Bregler, C.: Finding Pictures of Objects in Large Collections of Images. Object Representation in Computer Vision (1996) 335-360 108

[8] Gervautz, M., Purgathofer, W.: A simple method for color quantization: Octree quantization. Graphics Gems I (1990) 287-293 111

[9] Haralick, R. M., Shapiro, L. G.: Image Segmentation Techniques. CVGIP Volume 29, Number 1 (January 1985) 100-132 109

[10] Huang, P. W., Jean, Y. R.: Spatial Reasoning And Similarity Retrieval For Image Database Systems Based On Rs-Strings. PR Volume 29, (1996) 2103-2114 113

[11] Huang, T., Rui, Y.: Image retrieval: Past, present, and future. International Symposium on Multimedia Information Processing, (1997) 108

[12] Karu, K., Jain, A. K., Bolle, R. M.: Is There Any Texture in the Image. ICPR96 (1996) B94.3 111

[13] Muñoz, X.: Image Segmentation Integrating Color, Texture and Boundary Information. Master Thesis. Universitat de Girona (2001) 109

[14] Shearer, K., Bunke, H., Venkatesh, S.: Video indexing and similarity retrieval by largest common subgraph detection using decision trees. PR Volume 34, Number 5 (May 2001) 1075-1091 109