CONTENT-BASED REPRESENTATION OF COLOUR IMAGE SEQUENCES

Ioannis Kompatsiaris and Michael G. Strintzis

Informatics and Telematics Institute
1, Kyvernidou Str.
54639 Thessaloniki, Greece
Email: ikom@iti.gr

ABSTRACT

In this paper, a procedure is described for the spatiotemporal segmentation and tracking of objects in colour image sequences. For this purpose, we propose the novel procedure of K-Means with connectivity constraint algorithm as a general segmentation algorithm combining several types of information including colour, motion and compactness. A new colour distance is also defined for this algorithm. The regularisation parameters are evaluated automatically using the min-max criterion. In this algorithm, the use of *spatiotemporal regions* is introduced since a number of frames is analyzed simultaneously and as a result the same region is present in consequent frames. Experimental results on real and synthetic colour data demonstrate the performance of the data.

1. INTRODUCTION

Digital video is an integral part of many newly emerging multimedia applications. New video coding standards, such as MPEG-4 and MPEG-7, do not concentrate only on efficient compression methods but also on providing better ways to represent, integrate and exchange visual information [1]. These efforts aim to provide the user with greater flexibility for "content-based" access and manipulation of multimedia data. Although the standards will provide the needed functionalities in order to compose, manipulate and transmit the "content-based" information, the production of these objects is out of the scope of the standards and is left to the content developer. Thus, the success of any content-based approach depends largely on the segmentation of the scene based on its image contents.

Segmentation methods for 2D images may be divided primarily into region-based and boundary-based methods [2, 3]. Region-based approaches [4] rely on the homogeneity of spatially localised features such as gray level intensity, texture and motion. Region-growing and split and merge techniques also belong to the same category. On the other hand, boundary-based methods use primarily gradient information to locate object boundaries. Segmentation of colour images has received significant attention from researchers because of the large amount of information contained in colour images. The clustering-based techniques for colour image segmentation normally choose the RGB space as the feature space. A few methods use colour spaces other than the RGB; for instance, the CIE $L^*a^*b^*$ colour space is preferred in [5].

The K-Means algorithm (also known as C-Means), originally devised by McQueen [6], is a classic clustering technique, which has in particular provided a solid basis for multispectral clustering. In this paper, a novel procedure for the segmentation of colour image sequences using both spatial and temporal information is presented. As a basis for the segmentation algorithm, the K-Means algorithm is used, modified so as to take into account the coherence of the regions. The regularisation parameters of this novel "K-Means with Connectivity Constraint" (KMC) algorithm are evaluated automatically using the min-max criterion [7]. The $L^*a^*b^*$ colour space is used, which is related to the CIE 1931 XYZ standard observer through a nonlinear transformation. The $L^*a^*b^*$ is a suitable choice because it is a perceptually equalised colour space, i.e. the numerical distance in this space is proportional to the perceived colour difference. The algorithm is extended so as to separate and track regions appearing in consequent frames of an image sequence. The methodology proposed here is similar to that proposed in [8] and thus differs significantly from that, most common in the literature, where an image is first separated into regions which are then tracked through time. In the proposed approach, a number of consequent frames of the image sequence are analyzed *simultaneously* in order to segment the images into regions. This *higher order segmentation* implicitly solves the problem of correspondence of objects between consequent frames.

The paper is organised as follows. In the following Section the K-Means with connectivity constraint is described.

This work was supported by the EU CEC Project INTERFACE (IST Project 10036) and the Greek Secretariat of Research and Technology Project PENED99.

In Section 3 the use of the min-max criterion is discussed for the automatic evaluation of the segmentation regularisation parameters. The final spatiotemporal segmentation algorithm is presented in Section 4. Experimental results evaluating the performance of the algorithm are given in Section 5.

2. THE K-MEANS WITH CONNECTIVITY CONSTRAINT ALGORITHM

Clustering based on the K-Means algorithm is a widely used region segmentation method [9] which, however tends to produce unconnected regions. This is due to the propensity of the classical K-Means algorithm to ignore spatial information about the intensity values in an image, since it only takes into account the global intensity or colour information. In order to alleviate this problem, we propose the use of an extended K-Means algorithm: the K-Means-with-connectivity-constraint algorithm. In this algorithm the *spatial proximity* of each region is also taken into account by defining a new spatial center for the K-Means algorithm. For this algorithm, as in [5], we use $L^*a^*b^*$ colour space for colour distance definition.

- Step 1 An initialisation procedure based on colour histogram analysis results in K initial regions, with colour centers $\bar{\mathbf{I}}_k$, where $\bar{\mathbf{I}}_k = (\bar{I}_{L,k}, \bar{I}_{a,k}, \bar{I}_{b,k})$, and spatial centers $\bar{\mathbf{S}}_k = (\bar{S}_{k,x}, \bar{S}_{k,y})$, $k = 1, \ldots, K$.
- Step 2 For every pixel $\mathbf{p} = (x, y)$ the colour differences are evaluated between center and pixel colours as well as the distances between $\mathbf{p}$ and $\bar{\mathbf{S}}_k$ and $\mathbf{v}(\mathbf{p})$ and $\bar{\mathbf{v}}_k$, where $\mathbf{v}(\mathbf{p})$ are the motion parameters of each pixel, estimated by a motion estimation procedure. The motion estimation strategy we adopted for this algorithm is based on traditional block motion estimation techniques [10]. A generalised distance of a pixel $\mathbf{p}$ from a region s_k is defined as follows:

  $$D(\mathbf{p}, k) = \frac{\lambda_1}{\sigma_I^2}\|\mathbf{I}(\mathbf{p}) - \bar{\mathbf{I}}_k\| + \frac{\lambda_2}{\sigma_V^2}\|\mathbf{v}(\mathbf{p}) - \bar{\mathbf{v}}_k\| + \frac{\lambda_3}{\sigma_S^2}\bar{A}\frac{\|\mathbf{p} - \bar{\mathbf{S}}_k\|}{A_k}$$

  where $\sigma_I, \sigma_V, \sigma_S$ are the standard deviations of colour, motion and spatial distance, respectively and $\lambda_1, \lambda_2, \lambda_3$ are regularisation parameters. The colour distance between a pixel and a colour center is defined as:

  $$\|\mathbf{I}(\mathbf{p}) - \bar{\mathbf{I}}_k\| = \sqrt{(I_L - \bar{I}_{L,k})^2 + (I_a - \bar{I}_{a,k})^2 + (I_b - \bar{I}_{b,k})^2}$$

  The area of each region A_k is defined by $A_k = M_k$, where M_k is the number of pixels assigned to s_k and the mean area of all regions $\bar{A} = \frac{1}{K}\sum_{k=1}^{K} A_k$. For the first iteration it is assumed that $\lambda_2 = 0$ and that $A_k = \bar{A}$. Normalisation of the spatial distance, $\|\mathbf{p} - \bar{\mathbf{S}}_k\|$ by division with the area of each region, $\frac{\bar{A}}{A_k}$ is used in order to encourage the creation of large connected regions; otherwise, pixels with similar colour and motion values with those of a large region may be assigned to separate neighbouring smaller regions. If $|D(\mathbf{p}, i)| < |D(\mathbf{p}, k)|$ for all $k \neq i$, $\mathbf{p} = (x, y)$ is assigned to region s_i.
- Step 3 Following the new subdivision, all centers are recalculated. If M_k elements are assigned to s_k then $\bar{\mathbf{I}}_k$ is

  $$\bar{\mathbf{I}}_k = \frac{1}{M_k}\sum_{m=1}^{M_k} \mathbf{I}(\mathbf{p}_m^k) \ , \tag{1}$$

  and

  $$\bar{S}_{k,x} = \frac{1}{M_k}\sum_{m=1}^{M_k} p_{m,x}^k, \bar{S}_{k,y} = \frac{1}{M_k}\sum_{m=1}^{M_k} p_{m,y}^k, \tag{2}$$

  where $\mathbf{p}^k = (p_x^k, p_y^k)$, and the motion centers $\bar{\mathbf{v}}_k = (\bar{v}_{k,x}, \bar{v}_{k,y})$ are defined by

  $$\bar{v}_{k,x} = \frac{1}{M_k}\sum_{m=1}^{M_k} v_x(\mathbf{p}_m^k), \bar{v}_{k,y} = \frac{1}{M_k}\sum_{m=1}^{M_k} v_y(\mathbf{p}_m^k) \ , \tag{3}$$
- Step 4 If the difference between the new and the old centers $\bar{\mathbf{I}}_l$, $\bar{\mathbf{S}}_l$ and $\bar{\mathbf{v}}_l$ is below a threshold then goto *Step 5*, else goto *Step 2* using the new colour, motion and spatial centers.
- Step 5 Due the connectivity constraint, a uniform region may be separated into a number of smaller regions. A very simple region merging algorithm suffices for the extraction of the final regions.

An example of the segmentation procedure is shown in Fig. 1. Fig. 1a shows the original image of the videoconference sequence “Claire” of size 176×144. Fig. 1b shows the result of the KM algorithm. The result of the KMC algorithm (Step 4 of the algorithm) is in Fig. 1c. Fig. 1d shows the final segmentation after only four iterations.

3. SELECTION OF THE REGULARISATION PARAMETERS

The segmentation procedure described in Section 2 produces a segmentation $\mathcal{S}^*$ by assigning each pixel of the foreground object to one of the L regions s_l, $l = 1, \ldots, L$. This segmentation minimises the following energy function [11]:

$$E(\mathcal{S}, \boldsymbol{\lambda}) = \sum_{l=1}^{L}\sum_{p \in s_l} \boldsymbol{\lambda}^T \mathbf{D}(\mathbf{p}, l), \tag{4}$$

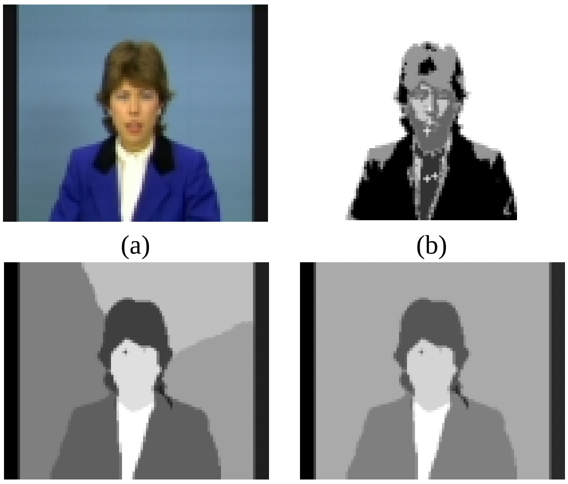

Fig. 1. (a) Original image "Claire". (b) Result of the KM algorithm. (c) Result of the KMC algorithm. (d) The final segmentation after only four iterations.

where

$$\boldsymbol{\lambda} = \begin{bmatrix} \frac{\lambda_1}{\sigma_I^2} \\ \frac{\lambda_2}{\sigma_V^2} \\ \frac{\lambda_3}{\sigma_S^2} \end{bmatrix}, \quad \mathbf{D}(\mathbf{p}, l) = \begin{bmatrix} \|\mathbf{I}(\mathbf{p}^l) - \bar{\mathbf{I}}_l\| \\ \|V(\mathbf{p}^l) - \bar{v}_l\| \\ \frac{A_l}{A}\|\mathbf{p}^l - \bar{\mathbf{S}}_l\| \end{bmatrix}.$$

Clearly, the final segmentation depends on the regularisation parameter $\boldsymbol{\lambda}$. In our case, setting $\lambda_1 \gg \lambda_2$ emphasises the importance of colour information and encourages the production of many unconnected regions. By contrast, when setting $\lambda_2 \gg \lambda_1$, regions with similar motion are created. In both cases, if $\lambda_1 + \lambda_2 \gg \lambda_3$ the segmentation results in a partition of L connected regions regardless of the colour or motion information. The weight factor $\boldsymbol{\lambda}$ must be constrained, because otherwise, as $\boldsymbol{\lambda}$ increases without limit, so does the energy $E(\mathcal{S}, \boldsymbol{\lambda})$. Thus we set $\|\boldsymbol{\lambda}\| = 1$.

This parameter may be chosen heuristically or by using a priori knowledge. Alternately, the *min-max criterion* [7] can be used for automatically evaluating the best $\boldsymbol{\lambda}$. This criterion is based on the assumption that in situations with conflicting alternatives, the most rational strategy is the one aiming to minimise the maximum possible losses. Thus, the function is minimised over all possible combinations of weight values and the min-max criterion selects the combination producing the maximum of these minima. In this case, the application of the min-max criterion gives:

$$\begin{array}{ll} E(\mathcal{S}^*, \boldsymbol{\lambda}^*) \geq E(\mathcal{S}^*, \boldsymbol{\lambda}) & \text{if } \|\boldsymbol{\lambda}\| = 1,\ \boldsymbol{\lambda} \neq \boldsymbol{\lambda}^* \\ E(\mathcal{S}^*, \boldsymbol{\lambda}^*) \leq E(\mathcal{S}, \boldsymbol{\lambda}^*) & \text{if } \mathcal{S} \neq \mathcal{S}^* \end{array}$$

4. SPATIOTEMPORAL SEGMENTATION

The segmentation procedure along with the selection of the regularisation parameters described above can be easily extended so as to separate and track regions appearing in consequent frames of an image sequence. The methodology proposed here differs from that, most common in the literature, where an image is first separated into regions and then these regions are tracked through time. In the proposed approach, a number of consequent frames of the image sequence are analyzed *simultaneously* in order to segment the images into regions.

If a small number T of consequent frames is used, it is reasonable to assume that the region colour remain substantially the same and hence that their colour centers $\bar{\mathbf{I}}_k$ are independent of time. In this way, region s_k is composed of all pixels $\mathbf{p}_t^k$ from T consequent frames. The KMC algorithm presented in Section 2 can be applied for a number of frames using the generalised distance $D(\mathbf{p}_t, k), t = 1, \ldots, T$ and the colour centers:

$$\bar{\mathbf{I}}_k = \frac{1}{M_k} \sum_{m=1}^{M_k} \mathbf{I}(\mathbf{p}_{m,t}^k),$$

where $\mathbf{p}_{m,t}^k$ are all pixels belonging to region s_k at time t. A similar assumption of independence of time for the spatial and motion centers cannot be made, since the region may move or move with varying velocity from frame to frame. Thus,

$$\bar{\mathbf{S}}_{k,t} = \frac{1}{M_k} \sum_{m=1}^{M_k} \mathbf{p}_{m,t}^k,$$

where only the pixels belonging to region k at frame t contribute to the estimation of the spatial center of subobject k at frame t. Similarly

$$\bar{\mathbf{v}}_{k,t} = \frac{1}{M_k} \sum_{m=1}^{M_k} \mathbf{v}(\mathbf{p}_{m,t}^k).$$

The energy measure (4) to be minimised becomes

$$E(\mathcal{S}) = \sum_{l=1}^{L} \sum_{p \in s_l} \boldsymbol{\lambda}^T \mathbf{D}(\mathbf{p}_t^l, l), \tag{5}$$

where

$$\mathbf{D}(\mathbf{p}_t^l, l) = \begin{bmatrix} \|\mathbf{I}(\mathbf{p}_t^l) - \bar{\mathbf{I}}_l\| \\ \|\mathbf{v}(\mathbf{p}_t^l) - \bar{\mathbf{v}}_{l,t}\| \\ \frac{A_l}{A}\|\mathbf{p}_t^l - \bar{\mathbf{S}}_{l,t}\| \end{bmatrix}.$$

A temporal component labeling algorithm is performed at Step 3 of the algorithm and the connectivity is now checked in three dimensions; x, y and time t. The algorithm produces connected regions appearing in consequent frames.

5. EXPERIMENTAL RESULTS

The algorithm described above was tested for the segmentation a synthetic colour test image (Fig. 2a) The synthetic image contains regions with very similar shades of the blue colour corrupted by Gaussian noise, making the automatic segmentation a challenging task. The algorithm was performed for one frame only, without any motion information in order to segment the static colour images. The segmentation result can be seen in Fig. 2b.

The algorithm described above was also used for the segmentation of the general image sequences "Table Tennis". The original image can be seen in Fig. 2c. The K-Means with connectivity constraint algorithm was applied in order to segment the sequences into meaningful regions. The segmentation results for the first three frames are shown in Fig. 2d-f. By merging regions with significant motion, moving objects are detected. The results for the first two frames are shown in Fig. 2g-h.

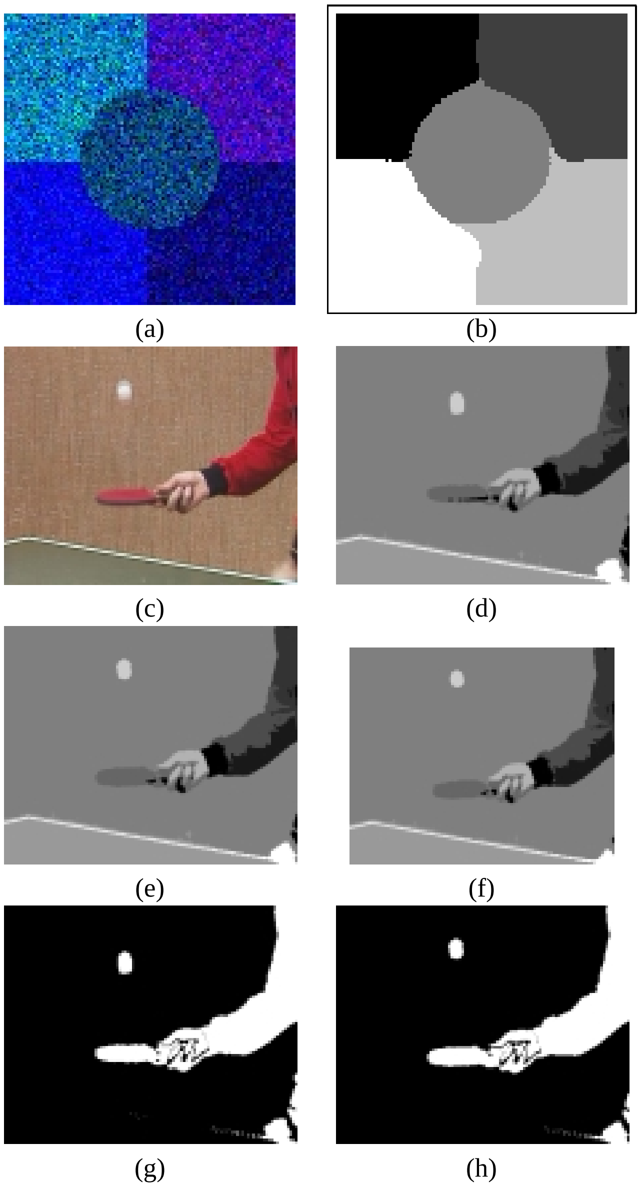

Fig. 2. (a) Synthetic test image. (b) Segmented image. (c) Original "Table Tennis" image. (d)-(f) Segmented frames. (g)-(h) Separation of moving objects.

6. REFERENCES

[1] R. Koenen, "MPEG-4 Multimedia for our Time," *IEEE Spectrum*, vol. 36, no. 2, pp. 26–33, Feb. 1999.

[2] "Special Issue," *IEEE Trans. on Circuits and Systems for Video Technology, Special Issue on Image and Video Processing*, vol. 8, no. 5, Sept. 1998.

[3] "Special Issue," *Signal Processing, Special Issue on Video Sequence Segmentation for Content-Based Processing and Manipulation*, vol. 66, no. 2, April 1998.

[4] P. Salembier and F. Marques, "Region-Based Representations of Image and Video: Segmentation Tools for Multimedia Services," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1147–1169, December 1999.

[5] S. Liapis, E. Sifakis, and G. Tziritas, "Color and/or Texture Segmentation using Deterministic Relaxation and Fast Marching Algorithms," in *Intern. Conf. on Pattern Recognition*, September 2000, vol. 3, pp. 621–624.

[6] J. McQueen, "Some Methods for Classification and Analyis of Multivariate Observations," in *5th Berkely Symp. on Math. Stat. and Prob.*, 1967, vol. 1, pp. 281–296.

[7] M. A. Gennert and A. L. Yuille, "Determining the Optimal Weights in Multiple Objective Function Optimization," in *2nd Int. Conf. on Comp. Vision*, 1988.

[8] S. Ayer, *Sequential and Competitive Methods for Estimation of Multiple Motions*, Ph.D. thesis, Ecole Polytechnique Federale de Lausanne, 1995.

[9] I. Kompatsiaris and M. G. Strintzis, "Spatiotemporal Segmentation and Tracking of Objects for Visualization of Videoconference Image Sequences," *IEEE Trans. Circuits and Systems for Video Technology*, to appear 2000.

[10] T. Sikora, "The MPEG-4 Video Standard Verification Model," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 19–31, Feb. 1997.

[11] R. Castagno, T. Ebrahimi, and M. Kunt, "Video Segmentation Based on Multiple Features for Interactive Multimedia Applications," *IEEE Trans. on Circuits and Systems for Video Technology, Special Issue on Image and Video Processing for Emerging Interactive Multimedia Services*, vol. 8, no. 5, pp. 562–571, Sept. 1998.