VIP: Vision tool for comparing Images of People

Michel Lantagne, Marc Parizeau and *Robert Bergevin*

Laboratoire de vision et systèmes numériques (LVSN),
Département de génie électrique et de génie informatique,
Université Laval, Ste-Foy (Qc), Canada, G1K 7P4.
E-mail: {lantagne, parizeau, bergevin}@gel.ulaval.ca

Abstract – *This paper describes the VIP technique, a Vision tool for comparing Images of People. This technique compares two human silhouettes and produces a similarity score between them. VIP was developed in the context of a surveillance project where one of the objectives is to recognize, in real-time, a person over different angles. The silhouette comparison must be robust to real-world situations, in particular to variations in scales, lighting conditions and human pose. The development of VIP involved the merging of several content-based image retrieval techniques. Colour and texture descriptors are used to describe the regions found inside a person's silhouette and a region matching scheme associates regions of one silhouette to another. For a recall of 80%, the average precision was found to be between 76% and 93% for a 870 images database with 16 different people. These results show the robustness of the system against scale, angle, and rotation variations in human silhouette appearance.*

Keywords: Similarity Measure, Feature, Descriptor, Colour, Texture, Region Matching,

1 Introduction

With the recent availability of cheap but powerful computer hardware, one can now envision the emergence of sophisticated and intelligent surveillance systems integrating a network of loosely-coupled computation nodes, each connected to a camera. These systems would need to track a person from non overlapping fields of view in order to determine whether the cameras observe the same person.

In this context, the VIP technique was developed to compare human silhouettes. A human silhouette refers to the contour of a person in an image and all the information contained within. This task is complex because important changes in the person's appearance can appear over different angles. The surveillance context imposes another requirement, it has to achieve real-time processing. Thus, the developed algorithms must be efficient.

Human silhouette comparison can be adressed by characterizing a person's appearance. For VIP, this characterization is conducted as follows. The first step is to divide an *a priori* extracted human silhouette into three parts. These three parts (upper, middle and lower parts) correspond respectively to the head, to the trunk and arms and to the legs. Then, each silhouette part is segmented into significant regions. The JSEG algorithm [4] is used for this automatic segmentation. The next step is to calculate descriptors of colour and texture for each region. The colour descriptor is a modified version of the descriptor presented in [5]. The texture descriptor is efficient and simple, and is based on those described in [10] and [7]. Then, a similarity measure between two regions is defined. And finally, to compare the regions inside two silhouette parts, a region matching scheme is used, involving a modified version of the IRM algorithm [10]. The output is a score between 0 and 1 which indicates the similarity between the two compared people, where value 1 corresponds to two identical silhouettes.

The article structure is as follow. Section 2 presents the context of CBIR systems and techniques. Section 3 describes the VIP technique. Section 4 presents experimental results. Finally Section 5 concludes the paper.

2 Related Work

The Content-Based Image Retrieval (CBIR) domain proposes several systems (such as [1, 8, 10]) and techniques (for example [3, 5, 7, 9]) to characterize the general appearance of an image or a region thereof.

In order to understand how a CBIR system works, it is necessary to consider the following two elements: features and descriptors. A feature is characteristic information that has a meaning for certain users or certain applications. The colour of a region in an image or the texture type such as directionality or repetition, are some examples. A descriptor is a model which assigns a value (or value set) for one or more features. A traditional example is the colour histogram.

A general description of a person's appearance for similarity-based retrieval must take into account the com-

bination of various feature descriptors. The descriptors most often used in CBIR systems are related to colour and texture properties.

Colour feature Colour is one of the most important visual feature. It is immediately perceived when looking at an image. The minimum structure of a colour description with a discrete system consists of a colour space definition, a quantization of the colour space, and a colour representation.

Texture feature Texture is another powerful discriminating feature, present almost everywhere in nature. Texture is a broad term used in pattern recognition to identify image patches that are characterized by differences in brightness. The texture of a visual item characterizes the interrelationship between adjacent pixels.

Similarity and distance To compare two descriptors, it is necessary to define a similarity (or distance) measure. The similarity measure defines an interval between 0 and 1, where value 1 corresponds to two identical descriptors. The distance measure returns the value 0 if the two descriptors are identical and an increasing value if the two descriptors are different.

2.1 CBIR systems

VIP is inspired by techniques found in two CBIR systems which have been developed in the past. Herein we present a general overview of the descriptors and region matching schemes used in these two systems.

Simplicity This system, developed by Wang *et al.* [10], uses semantic classification methods and a wavelet-based approach for feature extraction. An image is represented by a set of regions, which are characterized by colour, texture, shape, and location. The colour features are the averages in L, U, V components of colour, and the texture features are the energy of the wavelet coefficients.The shape features are normalized inertia of order 1 to 3. Simplicity uses IRM (Integrated Region Matching) as a region matching scheme. One region can be matched to one or more other regions. This way, the system is more robust to poor segmentation.

VisualSeek Smith and Chang [8] developed a system that supports retrieval of images based on colour and texture features. VisualSeek integrates feature-based image indexing with spatial query methods. The colour regions are represented by colour sets. Colour sets, different from histograms, are binary vectors that correspond to a selection of colours. The HSV colour model was used and a fixed quantization into 166 bins is made. The quantization gives 18 hues, 3 saturations, 3 values and 4 grays. Also VisualSeek uses the histogram quadratic distance metric [6] to compute distance colour between image regions.

3 VIP description

This section presents the VIP technique. The colour and texture descriptors are described in Subsections 3.1 and 3.2 respectively. The similarity measure between two regions is presented in Subsection 3.3 and the region matching scheme is described in Subsection 3.4. The global similarity measure for comparing human silhouettes is presented in Subsection 3.5.

3.1 Colour descriptor

VIP defines the *dominant colour descriptor* which takes into account the significant colours of the region. This descriptor is a version of the one described in [5]. The descriptor is based on the observation that a small number of colours is usually sufficient to characterize the colour information in an image region.

The descriptor presented in [5] uses a perceptual colour quantization algorithm [3] to obtain a small selection of colours for each region. Due to the high complexity of this clustering algorithm, we prefer to use a fixed quantification of the HSV space as described in [9], keeping in mind that VIP is aiming for real-time execution. HSV allows a colour analysis according to its more natural components: hue, saturation and brightness. A non-linear non-separable quantizer with 166 bins is used (18 hues $\times$ 3 saturations $\times$ 3 values + 4 grays = 166 colours). For a HSV colour $c = (h, s, v)$, $h \in [0°, 360°], s \in [0, 1]$ and $v \in [0, 1]$, the corresponding quantized colour $q = \{0, 1, ..., 165\}$ is obtained as follow.

$$q = \begin{cases} 0 & \text{if } v \leq 0.1 \\ g(h,s,v) & \text{if } s < 0.1 \text{ and } v > 0.1 \\ f(h,s,v) & \text{otherwise} \end{cases} \quad (1)$$

where

$$g(h,s,v) = \begin{cases} 1 & \text{if } s < 0.1 \text{ and } 0.1 < v \leq 0.4 \\ 2 & \text{if } s < 0.1 \text{ and } 0.4 < v \leq 0.7 \\ 3 & \text{if } s < 0.1 \text{ and } 0.7 < v \leq 1.0 \end{cases} \quad (2)$$

and $f(h, s, v)$ divides evenly the remaining HSV space with $20°$ intervals for hue and 0.3 intervals for saturation and value.

After the quantization step, only a small number of colours remain. The normalized colour histogram is calculated. The resulting histogram bins are now related to the percentage of a colour for a region. A threshold is applied on these percentages to keep only the significant colours. In this case, a colour is considered dominant if it covers more than five percent (5%) of the region area. The descriptor FC is then formed by the dominant pairs of colour and percentage:

$$FC = \{\{c_i, p_i\}, i = 1, ..., N, p_i \in [0, 1]\} \quad (3)$$

where N is the number of dominant colours of the region, c_i its colour and p_i its percentage.

To compare two FC descriptors, the quadratic colour histogram distance measure $D_h^2(H_1, H_2)$ for histograms H_1 and H_2 defined in [6] is used:

$$D_h^2(H_1, H_2) = (H_1 - H_2)^T A (H_1 - H_2) \quad (4)$$

where A is a matrix of weights that takes into account the cross correlation between histogram bins. Then, the distance $\mathrm{d}_c(FC_1, FC_2)$ between two descriptors $FC_1 = \{\{c_i, p_i\}, i = 1, ..., M\}$ and $FC_2 = \{\{c'_j, q_j\}, j = 1, ..., N\}$ is (ignoring all the zero entries):

$$\mathrm{d}_c(FC_1, FC_2) = \sum_{i=1}^{M}\sum_{k=1}^{M} a_{i,k} p_i p_k + \sum_{j=1}^{N}\sum_{l=1}^{N} a_{j,l} q_j q_l - \sum_{i=1}^{M}\sum_{j=1}^{N} 2 a_{i,j} p_i q_j \quad (5)$$

The coefficients $a_{i,j}$ of A represent the similarity between two colours:

$$a_{i,j} = 1 - d_{i,j}/d_{max} \quad (6)$$

where $d_{i,j}$ is the Euclidean distance between colours i and j, and d_{max} is the maximum distance between two colours. For two colours in HSV space, (h_i, s_i, v_i) and (h_j, s_j, v_j),

$$d_{i,j} = [(v_i - v_j)^2 + (s_i \cos h_i - s_j \cos h_j)^2 + (s_i \sin h_i - s_j \sin h_j)^2]^{\frac{1}{2}} \quad (7)$$

and $d_{max} = \sqrt{5}$ is obtained by computing two opposing colours, such as $(0°, 1, 0)$ and $(180°, 1, 1)$.

3.2 Texture descriptor

VIP defines a texture descriptor called *edge energy descriptor*. It is based on the idea for the texture descriptor in the Simplicity system [10]. This descriptor characterizes the edge density inside a region according to different orientations. The problem of scale is solved with a normalization of the density by the total number of pixels of the region. The idea behind this descriptor is the fact that intensity variations in a direction strike a human observer. The main advantage of the descriptor is that it is simple and fast to compute.

The first step is the conversion of the colour pixels to grayscale. Then, four edge detectors [7] are applied on the region. Horizontal ($0°$), vertical ($90°$) and diagonal ($35°$ and $135°$) edges are calculated. The four edge detectors are shown in Figure 1.

1	1	1	-1	$\sqrt{2}$	0	0	$\sqrt{2}$
-1	-1	1	-1	0	$-\sqrt{2}$	$-\sqrt{2}$	0

Figure 1: *Edge detectors.*

For a given orientation θ, the energy of the edges is calculated as follows:

$$E_\theta = \sqrt{\frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N} e_{\theta(i,j)}^2} \quad (8)$$

where $e_{\theta(i,j)}$ is an edge pixel at (i, j) in the region and MN is the number of edge pixels in that region.

The edge energies form the descriptor FT:

$$FT = \{\{E_\theta\}, \theta = 0°, 45°, 90°, 135°\}. \quad (9)$$

To compare two *edge energy descriptors*, a distance measure is defined. The distance $D_t(FT_1, FT_2)$ between two descriptors $FT_1 = \{\{E_\theta\}, \theta = 0°, 45°, 90°, 135°\}$ and $FT_2 = \{\{E'_\theta\}, \theta = 0°, 45°, 90°, 135°\}$ is:

$$D_t(FT_1, FT_2) = \sum_{\theta}(E_\theta - E'_{\theta+\phi})^2 \quad (10)$$

where ϕ represents the correspondence between the energies of FT_1 and FT_2. This way, the distance becomes relatively invariant to the rotation of the region in the image.

The angle ϕ is calculated as the difference between orientation θ_1 of greatest energy of FT_1 and orientation θ_2 of greatest energy of FT_2. Thus, the two strongest orientations are compared together and the other orientations are compared according to rotation. For example, let $\theta_1 = 90°$ and $\theta_2 = 45°$ be the orientations, respectively, for the greatest energy of region 1 and 2. Then, $\phi = 45° - 90° = -45°$. Therefore, $E_{90°}$ is compared with $E'_{45°}$, $E_{135°}$ with $E'_{90°}$ and so on. Note that, to be consistent in (10), $180°$ is added to ϕ if $(\theta + \phi) < 0°$, and $180°$ is removed if $(\theta + \phi) > 135°$.

3.3 Similarity measure

The similarity measure between two regions combines the colour and texture descriptors. First, two region sets, $A = \{a_1, a_2, ..., a_m\}$ and $B = \{b_1, b_2, ..., b_n\}$, are defined. For a short notation, the similarity between regions a_i and b_j for the colour descriptor is noted $s_c(a_i, b_j) = s_c(FC_{a_i}, FC_{b_j})$ and for the texture descriptor, $s_t(a_i, b_j) = s_t(FT_{a_i}, FT_{b_j})$. The similarity $s(a_i, b_j)$

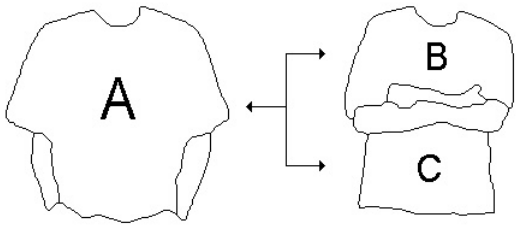

Figure 2: *Example that shows two segmentations of the shirt of the same person in two different poses. The result is a one-region shirt (A) for the first image and a two-region shirt (B+C) for the second.*

between two regions is the weighted sum of the two descriptor similarities:

$$s(a_i, b_j) = \alpha \cdot s_c(a_i, b_j) + (1 - \alpha) \cdot s_t(a_i, b_j) \quad (11)$$

where the α parameter represents the relative importance of the two descriptors. The use of this parameter will be discussed later.

The colour similarity s_c and the texture similarity s_t are computed directly from distances d_c and d_t (earlier defined in Subsections 3.1 and 3.2) as used in [1]:

$$s_c(a_i, b_j) = \exp(-\mathrm{d}_c(a_i, b_j)/\sigma_c) \quad (12)$$

$$s_t(a_i, b_j) = \exp(-\mathrm{d}_t(a_i, b_j)/\sigma_t) \quad (13)$$

where σ_c and σ_t are the standard deviations of the distances computed over all of the database regions. The normalization of colour and texture similarities is important to apply the α parameter in (11).

3.4 Region matching scheme

The region matching scheme used in VIP is a modified version of the IRM (Integrated Region Matching) algorithm [10] which uses a similarity measure instead of distance (see previous Subsection 3.3). The advantage of IRM is the robustness against region segmentation results. Figure 2 illustrates this advantage. A region of a set can be matched with one or more regions of another set. For the global similarity, IRM integrates the properties of all the regions of the two sets.

The IRM algorithm is simple and works as follows. The global similarity $S(A, B)$ between two sets of regions A and B is the weighted sum of the similarities $s(a_i, b_j)$, $i = 1, ..., m$ and $j = 1, ..., n$ between their regions:

$$S(A, B) = \sum_{i,j} w(a_i, b_j) \cdot s(a_i, b_j) \quad (14)$$

where $w(a_i, b_j)$ is the weight between region a_i and b_j.

The first step is to calculate all of the similarities $s(a_i, b_j)$ as in (11). The principle of matching is to always match the most similar region pair first. Therefore, the similarities are sorted in decreasing order. In this manner, the first value of similarity corresponds to the best match between a region of A and a region of B. The second value corresponds to the second best match and so on.

The next step is the comparison of region areas. For each similarity $s(a_i, b_j)$ in decreasing order, percentage areas of regions a_i and b_j, over the total area of their corresponding region set, are compared. The weight $w(a_i, b_j)$ is set to the smallest percentage area between region a_i and b_j. The weight $w(a_i, b_j)$ represents the percentage of the two region sets A and B associated with the similarity $s(a_i, b_j)$.

Then, the percentage area of the largest region is updated by removing the percentage area of the smallest region so that it can be matched again. The smallest region will not be matched anymore with any other region. If two regions have the same percentage area, the weight is set to this percentage area, and the two regions are matched and removed from the process.

The process continues in decreasing order for all of the similarities $s(a_i, b_j)$. At each step, the largest area is updated. The global similarity $S(A, B)$ is the sum of all of the weights $w(a_i, b_j)$ and similarities $s(a_i, b_j)$ as in (14).

3.5 Human similarity measure

Having defined the similarity for two region sets, the global similarity between two silhouettes must now be defined. VIP considers three body parts in a silhouette: the first for the head, the second for the trunk and arms, and the last for the legs. Each part is a region set. To compute the global similarity between two silhouettes, VIP compares each of the three body parts. Let S_H, S_T, and S_L be the similarities respectively between upper (head), middle (trunk and arms) and lower parts (legs). Thus, the global similarity $S(P_A, P_B)$ between two human silhouettes P_A and P_B is the weighting sum between parts:

$$S(P_A, P_B) = 0.2 \cdot S_H + 0.5 \cdot S_T + 0.3 \cdot S_L \quad (15)$$

where the weights are chosen according to the average area of the parts over a normal human body.

4 Experiments

To evaluate VIP, images were acquired as follows. Figure 3 shows the sixteen actors (sixteen different people), known as PID 01 to PID 16. No clothing constraint was imposed; the people were accepted for the video sequences with the clothes they were wearing at the time of

Figure 3: *The sixteen humans for video sequence acquisitions. The video sequences are taken by three cameras adjusted to different apertures to simulate three lighting conditions. The walk of each person in the scene allows retrieval of several silhouettes at different scales and poses.*

Figure 4: *Images that illustrated the three simulated intensity levels for video sequence acquisitions.*

Figure 5: *Images that illustrated different scales and human poses.*

acquisition. Video sequences were taken by three cameras with the same view angle. The camera's apertures were adjusted to obtain three distinct levels of intensity, as illustrated in Figure 4. Three video sequences (one by each camera) is taken for each person. Each person follows a similar path, walking through the scene in many directions, to enable the retrieval of images of the same person at different scales and angles (see Figure 5).

For image selection, a video frame of the sequence is conserved at every 25 frames. Then, only images with a complete silhouette are kept. Images obtained by the first camera form the first database DB_1. The second image database DB_2 is formed with images taken by the second camera, and the third database DB_3 is composed of images taken by the third camera. Table 1 shows the number of images for each person and each database.

Silhouettes are extracted automatically by a simple background subtraction algorithm [11]. To facilitate this process, the scene is delimited by three white curtains. A low-complexity scene is justified because the silhouette extraction is not a part of VIP which objective is to compare human silhouettes, not to perform silhouette extraction.

Then, each extracted silhouette is divided into three parts: the upper part for the head, the middle part for the trunk and the arms, and the lower part for the legs. Currently, these three body parts are determined with a simple algorithm which assumes that people are standing. The upper part of the silhouette (head) refers to the upper 15% of the silhouette's height. The lower part (legs) refers to the lower 35% of the silhouette's height. The middle part (trunk and arms), corresponds to the remaining area of the silhouette.

Finally, the last initialization step is the segmentation of the three parts into regions. An automatic segmentation according to colour and texture is carried out using the JSEG [4] algorithm. Computation of colour and texture descriptors is performed for each region. The databases are now ready to perform evaluation of the technique.

PID	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	Total
DB_1	12	19	18	19	17	16	20	20	20	17	19	16	20	21	16	20	**290**
DB_2	18	19	21	19	16	18	21	19	21	16	18	16	20	21	15	22	**300**
DB_3	17	19	20	17	14	16	19	19	20	15	15	14	21	19	15	20	**280**
Total	47	57	59	55	47	50	60	58	61	48	52	46	61	61	46	62	**870**

Table 1: *Number of images for each person in the three databases. DB_1 regroups images taken by the first camera, DB_2, those taken by the second camera and DB_3, those taken by the third camera. The intensity level degrades from DB_1 to DB_3.*

4.1 Results

The query-by-example model was used in the experiments. A person is selected (the query) and VIP is used to compute similarities between the query and all other silhouettes in the selected database. For all experiments, the α parameter is set to 70% to devote more importance to the colour information. An interface shows images of people corresponding to similarity results ranking in decreasing order. As expected, the first rank is always occupied by the query. Two examples are shown in Figures 8 and 9. Note that the interface normalizes the silhouettes to display them in one size vignettes.

The quantitative measure used for evaluating performance for the experiments is the well known recall-precision measure issued from information retrieval [2]. The goal is to achieve a precision of 100% from a recall of 100%, but in practice this is not always possible for large databases. In general, higher precision leads to lower recall and vice versa.

To evaluate VIP results objectively, for an image query, the precision is computed from different recall values (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%). A recall of 100% indicates that all images of a person are to be retrieved. As mentioned before, all images of the database are sorted in decreasing order of similarity. Then, the precision can be defined as a function of the recall as follows:

$$P = \frac{R \times N(q)}{N_0} \tag{16}$$

where $N(q)$ is the total number of images of a query person q and N_0 is the total number of images needed to find $[R \times N(q)]$ images of that query person q. A precision of 100% indicates that the first $[R \times N(q)]$ results in the ranked list do not contain images of any other people.

For example, let person PID 07 of DB_1 be the query. VIP computes all similarities between the query and each of the 290 images in DB_1 and sorts them in decreasing order. PID 07 has 20 images in the database. A recall value of 100% corresponds to retrieving all 20 images of PID 07. If 25 images are necessary to obtain the recall value, which means that five false detections were observed, thus the precision is $20/25 = 80\%$.

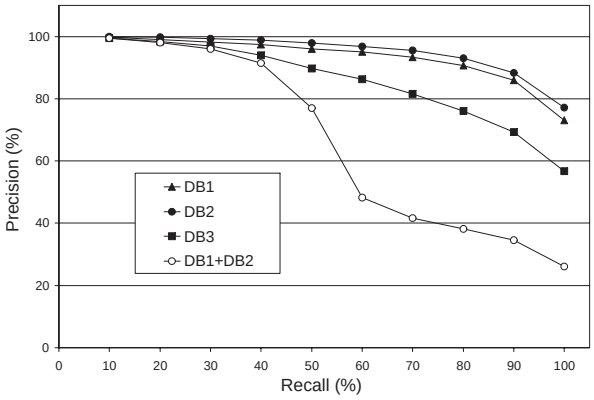

Figure 6: *Overall performance of VIP for the three databases. The* precision vs recall *curves are mean curves computed for all people.*

So, the resulting precision for one image query is interesting, but to analyse the behaviour of the technique for a distinct person, it is better to compute the mean *precision vs recall* curve for results of each image of that person. This way, the resulting curve can show the performance of VIP for a particular person. Then, a curve is obtained for each person in the database. Finally, the average curve of all individual curves demonstrates the overall performance of VIP for the database. This way, the fact that the precision can be negatively affected due to the ranking of one image is attenuated.

Having defined the tools to evaluate the performance, the technique's performance itself will now be examined. Figure 6 shows the results obtained with the three databases individually and for a mixture of DB_1 and DB_2 (two different intensities). VIP performs well with DB_1 and DB_2 individually with precisions of 91% and 93% for a recall of 80%. These results demonstrate the ability of VIP for comparing silhouettes of different scales and poses. The performance degrades with DB_3 with a precision of 80% for a recall of 70%. The explication is probably the too low level of image intensity. The results obtained for the combination of DB_1 and DB_2

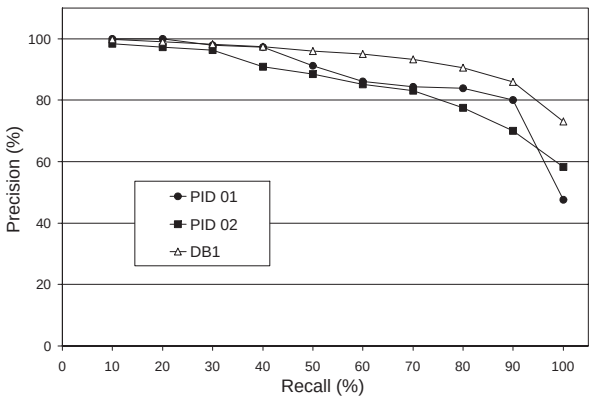

Figure 7: *The* precision vs recall *curves for PID 01 and PID 09 compared to DB_1 mean curve. PID 01 and PID 09 have similar clothes with respect to colour and texture.*

show that VIP was not robust enough, for the moment, to compare images with different intensities. The curve drops to a precision around 50% for a recall value of 60%.

Figure 7 shows the results for PID 01 and 09 with DB_1 (see Figure 3). The distance of these two curves with the average curve of DB_1 can be explained as follows. Images of PID 01 and PID 09 are composed of similar colours and region areas. Since VIP takes in account the colour and texture of regions found inside the silhouette and not the form, this is why the technique has considered PID 01 and PID 09 to be the same person.

5 Conclusions and Future Work

In this paper, the VIP technique was presented. VIP uses colour and texture features to compare two human silhouettes. A modified version of the *dominant colour descriptor* defined in [5] is used. This descriptor takes into account the significant colours of a region. VIP defines a texture descriptor called *edge energy descriptor*. This descriptor characterizes the edge density inside a region according to different orientations. To compare two sets of regions, a modified version of the region matching scheme IRM [10] is used.

The results presented show the accuracy of the system for the comparison of people. The precision from different recall values is used to illustrate the results. The first experiment on three databases of about 300 images of people shows the potential of VIP. The databases were composed of images of varying scales, lighting conditions and human pose. In the future, experiments will be performed on a larger database with more people wearing a greater variety of clothes.

References

[1] S. Ardizzoni, I. Bartolini, and M. Patella. Windsurf: Region-based image retrieval using wavelets. In *DEXA Workshop*, pages 167–173, 1999.

[2] Alberto Del Bimbo. *Visual information retrieval*. Morgan Kaufmann Publishers Inc., 1999.

[3] Yining Deng, Charles Kenney, Michael S. Moore, and B.S. Manjunath. Peer group filtering and perceptual color image quantization. In *Proceedings IEEE International Symposium on Circuits and Systems*, volume 4, pages 21–24, 1999.

[4] Yining Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):800–810, August 2001.

[5] Yining Deng, B. S. Manjunath, Charles Kenney, Michael S. Moore, and Hyundoo Shin. An efficient color representation for image retrieval. *IEEE Transactions on Image Processing*, 10(1):140–147, January 2001.

[6] J. Hafner, H.S. Sawhney, W. Equitz, M. Flickner, and W. Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736, July 1995.

[7] Dong Kwon Park, Yoon Seok Jeon, and Chee Sun Won. Efficient use of local edge histogram descriptor. In *Proceedings of the 2000 ACM workshops on Multimedia*, pages 51–54. ACM Press, 2000.

[8] John R. Smith and Shih-Fu Chang. Visualseek: a fully automated content-based image query system. In *Proceedings of the fourth ACM international conference on Multimedia*, pages 87–98. ACM Press, 1996.

[9] J.R. Smith. *Image Databases : Search and Retrieval of Digital Imagery*, chapter 11 - Color for Image Retrieval, pages 285–311. Wiley Inter-Science, 2002. V. Castelli and L.D. Bergman (Eds) - ISBN: 0-471-32116-8.

[10] James Z. Wang, Jia Li, and Gio Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947–963, September 2001.

[11] C. R. Wren, Ali Azarbayejani, Trevor Darrell, and Alex Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.

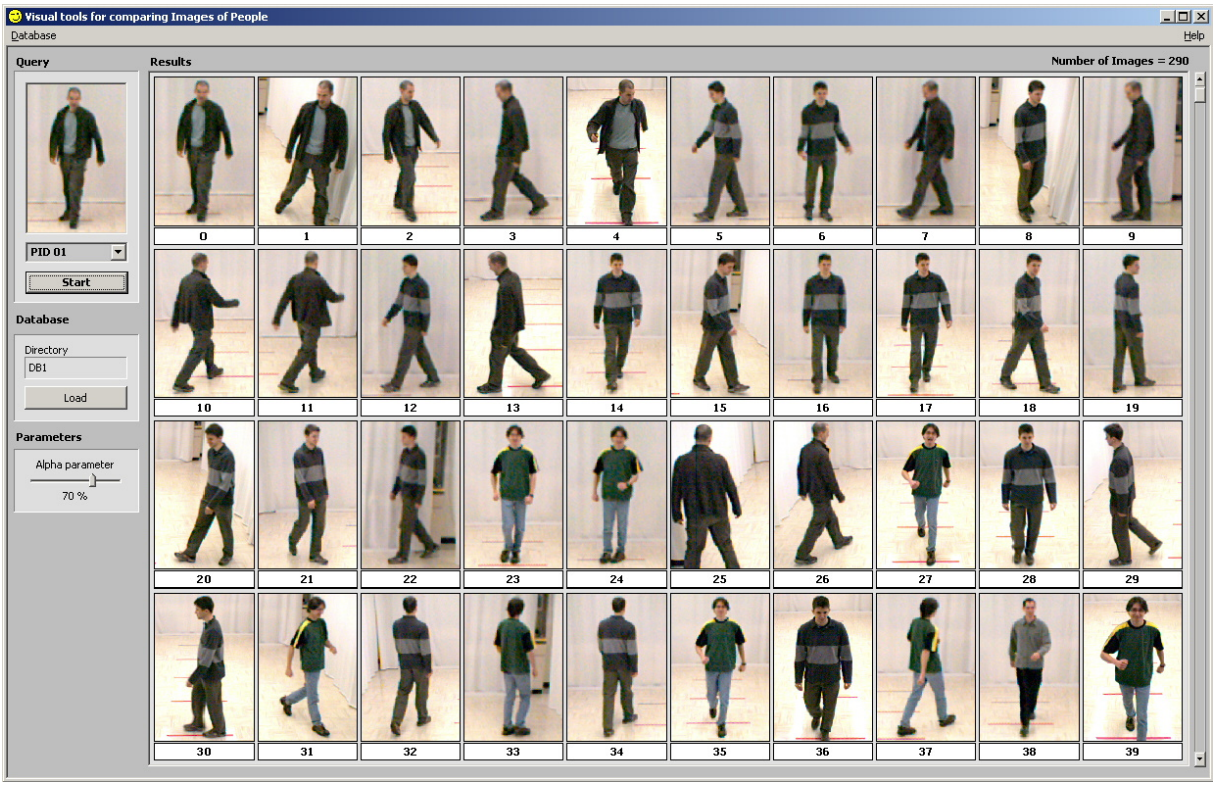

Figure 8: *Interface that shows results for PID 01 query of DB1 with parameter* $\alpha = 0.70$ *(silhouettes are normalized).*

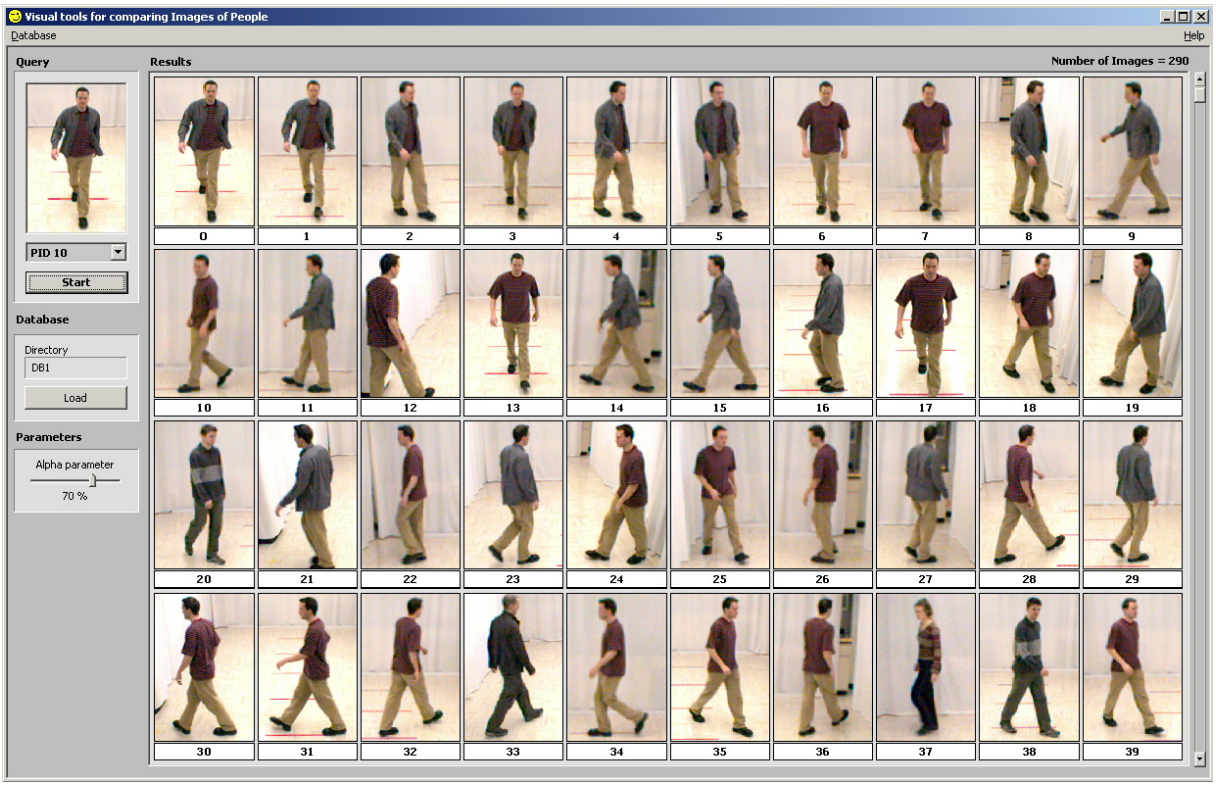

Figure 9: *Interface that shows results for PID 10 query of DB1 with parameter* $\alpha = 0.70$ *(silhouettes are normalized).*