IST 2006 – International Workshop on Imaging
Systems and Techniques
Minori, Italy 29 April 2006

Comparative Study of Adaptive Segmentation Techniques for Gesture Analysis in Unconstrained Environments

Martin Côté[1], Pierre Payeur[1], Gilles Comeau[2]

[1]School of Information Technology and Engineering
[2]Department of Music
University of Ottawa
Ottawa, Ontario, Canada, K1N 6N5
[mcote, ppayeur]@site.uottawa.ca
gcomeau@uottawa.ca

Abstract *– This paper discusses the importance of providing non-invasive techniques in the analysis of the complex movements performed by musicians and athletes. Current gesture analysis systems are insufficient and do not succeed in providing quality performance measurements without imposing strict environmental and operational constraints on individuals. Computer vision offers the means with which such techniques can be made possible without impeding the performance of musicians or compromising the integrity of the measurements. The following study compares some of the modern image segmentation techniques and discusses their shortcomings with respect to the stated context. A novel statistical method is also introduced in an attempt to improve the resilience of vision-based gesture segmentation.*

Keywords *– Adaptive image segmentation, tracking, statistical object recognition, gesture analysis, piano-playing evaluation.*

I. INTRODUCTION

With the advent of more powerful computing facilities and advances in artificial intelligence, few techniques have been proposed in the evaluation of the complex movements inherent in the performances of musicians and athletes. Such techniques are highly desirable in order to detect problematic situations which often lead to chronic injuries and to provide the capacity with which professors and trainers may measure the evolution and habits of individuals. Gesture analysis provides the means with which quantitative measurements may be obtained with respect to the physical performances of musicians and athletes.

Modern gesture analysis methodologies employed by professionals fail to obtain an exhaustive evaluation or a complete comparison between gestures. Current techniques in this type of analysis are not robust enough to operate in the unconstrained environments in which these performances must be evaluated. Many techniques today still rely on the use of encumbering sensor technologies and require the use of cabling or attaching markers on an individual. The techniques also often require a very controlled environment and involve tedious, expensive and complex operating methodologies. The imposed environments are usually foreign to the musicians or athletes resulting in an impeded performance and corruption to the exactitude of the measurements.

The undertaken study aims to research and develop new vision-based methodologies suitable for the analysis of musician gestures. Specifically the context being explored consists of analyzing the gestures and postures of pianists with applications to piano pedagogy. Computer vision techniques offer the ideal means for this type of evaluation. Due to their non-invasive nature they do not interfere with the performance of pianists. Computer vision techniques are put to the test in a context where the imposed constraints on the environment and individuals must remain minimal.

Image segmentation is an initial and important step towards a more comprehensive gesture analysis. It is often the first low-level process applied to images prior to a more complex high-level analysis. The goal of segmentation is to decompose an image into coherent sections that are meaningful to certain applications [1]. Recently these applications have included the extraction of video objects for modern video encoding formats [5, 9, 11, 13, 14], the indexing and retrieval of visual information [17, 19, 30, 32] and the analysis of human gestures [10, 18, 23, 29]. While multiple papers have already been proposed for this last application, many of them impose unreasonable constraints on the environment or do not adapt to possible changes in the scene.

This paper is organized as follows. Section II begins by introducing some of the common problems found in the segmentation of piano-playing sequences while maintaining minimal constraints on both the environment and the individual in order to conserve the exactitude of performance measurements. Section III looks at some of the more recent advances in image segmentation as well as their applicability within the stated context. Particular attention is given to statistical approaches, and in Section IV a technique combining the Mean-Shift Algorithm [16] with Multidimensional Receptive Field Histograms [31] is proposed. Section V presents some experimental results. Sections VI and VII offer a discussion on future work and a conclusion respectively.

II. CHALLENGES

In order for the proposed technique to be both useful and practical, the vision-based methodology employed must be able to adapt to the environment in which it is used. In the context of piano pedagogy the system must be able to handle various locations such as practice rooms and piano studios.

This means the system should have the capacity to operate in arbitrarily complex scenes where the background may contain any combination of stationary or non-stationary objects with varying colours and shapes. In such an environment, lighting conditions may be subject to changes and introduce persistent shadows.

As mentioned before, in order to preserve an objective measurement, pianists should not be required to conform to any type of constraints. Such flexibility implies that gestures must be segmented and tracked without imposing either a dress code or the presence of sensors on the pianists. The complex posturing of musicians means that the investigated techniques must be able to handle non-rigidity and a large range of possible motions. The success of the analysis should not rely on the use of sophisticated or expensive equipment but rather on the level of algorithmic performance.

III. State-of-the-Art Adaptive Segmentation

A. Contour-Based Approaches

An important class of segmentation techniques include those that rely on image edge information in order to delineate objects. One of the most popular edge-based techniques are Snakes (or active contours), introduced by Kass *et al*. [2]. The goal of this technique is to deform a contour so that it matches the boundary of a given object. The deformation of the contour is driven by an energy minimization procedure. The energy function involved is designed in such a way that its local minima are achieved when the contour corresponds to the boundary of the object. The technique has also been extended for the segmentation of video objects in [3]. In this case, the contours are projected into subsequent frames and re-adapted to the edge information of an object. The computational complexity is quite high and the technique in this case is not suited for large non-rigid movements.

Active contours are typically not appropriate for situations where objects may be partially occluded or where reliable edge-information is difficult to obtain. The former problem was resolved by Peterfreund [4] with the introduction of Kalman Snakes. Using a combination of optical flow measurements along with Kalman filtering, the contours were made resilient to partial occlusion effects. In the case of unreliable edge-information, Sun *et al.* [5] proposed the use of a Veterbi search algorithm to find the best possible positioning of key points along the contour.

B. Neural Network Approaches

Over the years, several techniques have proposed the use of neural networks for assisting in the segmentation process. Neural networks employ a large number of interconnected processing nodes that perform simple computations. Neural networks aim to imitate the biological reasoning capabilities of human beings. Their ability to learn and generalize patterns makes them powerful classifiers.

Several authors [6, 7, 8, 9] have proposed the used of Multi-Layered Perceptrons (MLPs) in order to classify image pixels into appropriate segmentation classes (typically foreground and background). In each case the networks are trained using a set of pre-segmented images that may or may not contain the object of interest. The major variations between techniques involve the choice of image features fed to the input layer of the network. [8] uses a simple three-dimensional RGB vector for classification, while [6] and [7] suggest the use of more comprehensive input vectors of 9 and 31 dimensions respectively. The networks' ability to segment is highly dependent on the available training data and the possible transformations an image may undergo.

The previously mentioned neural network classifiers are not however resilient to any kind of environmental changes. The network configurations remain static and do not adapt to changes in data representation. Only recently have a few papers proposed adding an adaptability mechanism to neural networks [9, 10]. These techniques rely on complex retraining algorithms in order to modify the network in response to a decrease in performance. Along with retraining the network, comes the difficulty in obtaining new training data and evaluating current segmentation performance.

C. Region-Based Approaches

Another popular approach to image segmentation consists of cutting an image into coherent regions that could be used to represent a given object. These approaches rely on the over-segmentation of an image in order to produce a set of perceptually homogenous regions. The two most common techniques for over-segmentation are the Watershed Transform [11] and the K-means clustering [12, 13] algorithm. The former uses image gradient information as a basis for segmentation. Each of the local minima found in an image produces a marker region corresponding to a perceptually uniform section. In the case of the K-means clustering algorithm pixels are grouped together in order to minimize the overall point-to-center distance for each cluster. The disadvantage of using a K-means clustering algorithm is that an initial estimation of the cluster centers is required. Furthermore, the K-means algorithm must be adapted to enforce a connectivity constraint in order to produce spatially consistent regions. Mezaris *et al.* [13] enforce connectivity by using a split-and-merge technique for disjoint regions while Pappas [12] adopts a Gibbs Random Field (GRF) representation for the image data.

Once over-segmentation has been achieved the next challenge involves finding region correspondences between frames. Mezaris *et al.* [13] proposed a trajectory tracking approach where regions are matched with their equivalents in subsequent frames. Tsai *et al.* [14] suggested merging regions along the time axis in order to formulate 3D volumes representing objects of interest. In the case of [15] it is the relation between region features that drives the tracking

process. They propose that an object's description can be enhanced by not only including the comprising regions, but also the manner in which they are related.

D. Statistical Approaches

Many techniques rely on the inherent statistical properties of the information found within an image in order to segment objects of interest. Three important statistical approaches are presented within this section. The first involves the use of a Gaussian distribution for the segmentation of human skin patches. The second uses Gaussian mixtures for background maintenance and finally the last approach is based on Comaniciu's [16] Mean-Shift Algorithm.

i. Probabilistic Skin Filters

Colour is an important and powerful cue in identifying objects, the advantage of using colour characteristics stem from the fact that they are highly invariant to most transformations [17]. That is, whether an object rotates, translates or deforms in some way, the colours that comprise it remain approximately the same. Many techniques have taken advantage of this fact and have used colour in the segmentation of human skin in order to segment faces and hands [17, 18, 19, 20].

While invariant to most movements, colours are highly susceptible to lighting conditions; shadows, sensor errors, light colour temperature and directionality of the light source, all contribute in the way colours are represented [20]. Yang *et al*. [21] have observed that despite environmental conditions, skin colours have a tendency to cluster within the RGB colour space. Specifically the clustering effect can be modeled using a Gaussian distribution. To mitigate the effects of illumination on the colour distributions, a normalization of the RGB tristimulus values can be done. Du *et al.* [18] have defined a 2D Gaussian probability function representing the likelihood of a pixel belonging to skin. A threshold is applied to the likelihood value to obtain an appropriate representation of skin patches within an image.

The challenge with this technique is in its initialization of the Gaussian distribution model. A sequence may potentially contain skin tones represented by different means and standard deviations. This requires that the system either has prior knowledge of the colours found within a sequence or that an operator initializes the model with a representative sample. The uniqueness of human skin colour is also a problem; non-human objects may often have near-skin colouration.

ii. Mixture of Gaussians

When segmenting individuals, a system must be able to identify more than just skin patches. Humans cannot adequately be represented using a single probabilistic model. In fact, for proper segmentation, a system must be able to identify a majority of the colours that make up the object of interest. Stauffer and Grimson [22] proposed a technique that uses a mixture of Gaussian probability models in order to identify the colours of an image and differentiate between the background and the foreground. This research has been used for many applications such as the tracking of human bodies [23] and the learning of patterns of activity in traffic monitoring [24].

The technique calls for a set of Gaussian distributions to be associated to each image pixel. This is in contrast to the preceding technique which only used a single distribution for all skin patches in an image. The likelihood of each new colour sample for a given pixel is verified against its set of distributions. If a matching distribution is not found, a new model is added to the mixture for the given pixel. The overall mixture model is refined using an on-line Expectation-Maximization (EM) algorithm. The technique also provides a learning mechanism for distinguishing between foreground and background models using a set of weights. The assumption is made that background models correspond to the distributions that most often describe a pixel's state. Some more recent publications have also added considerations for shadow and brightness changes in an image [25, 26].

The use of mixture of Gaussians has several shortcomings for the type of application considered in this research work. Its capacity to learn background and foreground models relies heavily on the motion of foreground objects. When tracking musicians, movements can be quite subtle, hence the colours that make up the individual change rather infrequently and would ultimately be considered as background. This problem represents a serious challenge in the application of the technique. The next difficulty in applying this algorithm stems from its initialization procedure. While it is not necessary to initialize the system without foreground objects, not doing so would mean that the modelling time of the background would be increased. Newly revealed background sections would register as foreground since the system would not have spent the required time learning the particular distributions of these sections.

iii. Continuously Adaptive Mean-Shift (CAMSHIFT)

The shortcomings of the mixture of Gaussians approach are also characteristic of most background subtraction methods. It is an accepted fact that most low-level computer vision operations such as segmentation should be assisted by higher-level information [16]. Due to the complexity of the scene and the lack of constraints on the environment, background subtraction techniques are not suitable for tracking human forms. By way of operator assistance, it is possible to track semantic components of an image such as individual body parts using feature recognition approaches.

As seen with the use of Gaussian probabilities, image colours can be represented using probabilistic models. The Mean-Shift Technique introduced by Comaniciu and Meer [16] uses statistical information to converge a search window over distribution modes. The approach is iterative and

independent of the underlining distribution parameters. The Mean-Shift Algorithm has been implemented in systems used for tracking of generic objects [27] and detecting vessels [28].

Intel researchers [29] have modified the original mean-shift technique to track specific colour distributions and to adapt their non-parametric descriptions to account for inter-frame differences. This new technique was named the Continuously Adaptive Mean-Shift (CAMSHIFT). Once initialized, a search window uses the histogram model of a local image area in order to produce a probabilistic representation of the image. Using the Mean-Shift Algorithm the window converges upon the distribution's mode and computes its size and location based on image moments. The window tracks the distribution in subsequent frames starting at its last converged position. The assumption is made that distributions are subjected to only small inter-frame differences and that the search window would require only a few steps before converging.

The advantage of using such a technique is that it does not require a parametric representation of the colour probabilities. The technique converges on the modes of the colour histograms without making any assumptions regarding their distribution. The algorithm works best when faced with unimodal distributions, however when segmenting semantic objects, complex colour patterns described by multi-modal distributions may be encountered. When tracking these complex colour patterns the algorithm fails to converge adequately over the multiple modes and results in incorrect position and size computations. These errors accumulate throughout a sequence and may result in a tracking failure. This is a major disadvantage to the technique since by the end of a sequence the tracked distributions may no longer reflect the original objects.

IV. Enhanced Camshift Algorithm

In order to mitigate some of the issues mentioned with the CAMSHIFT technique that relies solely on colour information in order to track objects of interest, we propose an original combination of the traditional technique with an extended colour histogram representation known as Multidimensional Receptive Field Histograms (MRFHs). The resulting technique provides a more comprehensive description of the tracked distributions.

In most segmentation approaches, colours are analyzed and tracked using colour histogram techniques originally proposed by Swain and Ballard [30]. But Schiele and Crowley [31] have successfully expanded on these traditional histograms and introduced a new method for describing the local appearance of objects. Their technique involves the use of Multidimensional Receptive Field Histograms (MRFHs) that include the response of a vector of local linear neighbourhood operators. MRFHs have been successfully applied for the identification of brand names appearing in advertisements during football and formula-1 video streams [32]. MRFHs extend the traditional colour histograms and include the response from several operators such as Gabor filters, Laplacian operators and Gaussian derivatives. With its use of multiple classes of data, MRFHs provide a much more discriminating description of objects and mitigate the problem of multi-modal colour distributions.

The enhanced CAMSHIFT algorithm increases the discrimination of the tracked distribution with respect to surrounding objects. This discriminating factor reduces noise from nearby distributions and results in a better localization of objects. While this framework extension does not fully resolve the issue of complex colour distributions, it does alleviate the influence they may have on the convergence process.

Preliminary experiment results revealed that for simplistic objects having highly discriminatory receptive field histograms, the enhanced CAMSHIFT probability map provided a very good approximation of the object of interest with low background noise. These approximations enable the implementation to discard the geometric search windows in favour of non-parametric shapes. Using a morphological dilation process followed with the application of a component size threshold it is possible to obtain an accurate segmentation of the tracked object. The newly acquired non-geometric shape of the object then serves as a more appropriate basis for the construction of the MRFH for subsequent frame segmentation.

The main limitation of using MRFHs for the segmentation of musicians stems from its inability to adequately describe some of the more varied objects. Some of the human sections being segmented have very complex colour features, vary greatly between frames and do not provide strong discriminating data. In these cases, larger MRFHs are required in order to capture more information and attempt to provide an accurate description; this leads to hefty memory requirements. With the abundance of information available it is also highly unlikely that a totally invariant description of these objects be obtainable. Another issue arises from the algorithm's inability to track less prominent image features, for example the pianist's hands. The objects in this case do not produce sufficiently dense histograms in order to be discriminating or accurately represent the object.

V. Experimental Results

For this experimental study, various statistical approaches where implemented and tested on video sequences of a musician playing the piano; no restrictions were placed on the environment or the individual. Figure 1a) shows an arbitrary frame of a test video sequence. Lighting is non-uniform throughout the scene, there are an abundance of shadows, light colour is modified by reflections and some of the musician's features are occluded due to the camera angle as well as loose fitting clothes. This sequence was selected due to its inherent complexity and serves as a good basis for discriminating between algorithms.

Figure 1b) demonstrates the results provided by the probabilistic skin filter. Due to the lack of exposed skin within the sequence obtaining a representative colour sample proved to be a difficult task. The change in lighting colour near the piano can be observed on the musician's hand; the skin colour in this area does not conform to the computed probability model. The filtering also produces a significant amount of noise around objects whose colours are close to the selected skin tone. While the technique could be expanded to include colour distributions other than skin, the noise level would also be significantly increased.

In the case of the mixture of Gaussians, multiple colour distributions are associated to each pixel in an attempt to classify between background and foreground. The algorithm has a hard time modeling the background when initialized with an object in the scene. As shown in Figure 1c), the wall portion located behind the musician is classified as foreground because it is a newly uncovered region where the algorithm has yet to learn colour distributions. On the other hand, certain portions of the musician are classified as background; these regions which are clearly part of the foreground are misclassified due to their lack of motion.

Figure 1d) demonstrates the probability map used by the CAMSHIFT algorithm in order to fit a geometric region (shown as a red rectangular window) over a tracked object; in this case, the musician's torso. Due to the simple histograms being used by the algorithm, a lot of noise is present within the mask. Regions of similar colour or having approximate characteristics are also partially identified. The musician's arm in this case has very similar colour characteristics as the torso; it introduces errors to the tracking of the object. The fitting of a parametric window for tracking objects has serious limitations; an individual is a deformable object, it cannot easily be modeled using simple geometry.

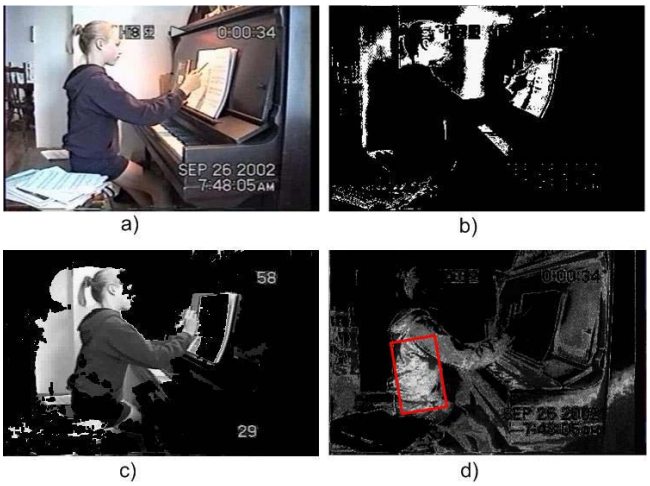

Figure 1 - a) Arbitrary video frame, b) Results of Probabilistic Skin Filter, c) Results of Mixture of Gaussians, d) Results of CAMSHIFT.

The final result shown in Figure 2 stems from the proposed enhancements on the CAMSHIFT algorithm. The approach was extended to use MRFHs and to allow non-geometric forms of the tracked objects. Details regarding the implementation are given in the previous section. The tracking of the torso and arm in this case is greatly improved over the other presented algorithms. A clear outline can be seen along the pianist's clothes giving a good indication of the body posture. The tracking of more complicated objects such as the musician's head yield results of a lower calibre as that of the torso. The complexity of this object hinders the system's ability to properly identify and track it. The head in this case has a very wide range in colouration and fails to produce a highly discriminating histogram. The hands, which cover a much smaller portion of the scene, do not yield a good segmentation due to their inability to produce sufficiently dense and invariant histograms.

Figure 2 - Results of Enhanced CAMSHIFT with MRFHs.

VI. Undergoing Developments

While the improved CAMSHIFT methodology presented here offers a good foundation for segmentation in unconstrained environments, more work needs to be done in order to produce better object descriptors. Even with the use of MRFHs, it is highly unlikely that every object of interest originating from arbitrarily complex scenes will have an adequately unique description. This problem is alleviated in other segmentation algorithms such as region-based approaches by having the algorithm supply the tracking regions.

This reveals the possibility of using hybrid segmentation techniques in order to overcome the limitations still encountered with the CAMSHIFT approach. Provided that a region-based procedure were able to offer both perceptually uniform and semantically coherent regions, the tracking procedure could be assisted by the statistical mechanisms presented within this paper. Alternatively, the over-segmentation process could be driven by an image's MRFH. This approach would result in the identification of regions with highly discriminating descriptors at the start of sequences.

VII. Conclusion

This paper presented a series of segmentation approaches and discussed their applicability to unconstrained environments as well their ability to adapt to complex inter-frame differences. A novel approach which combines the CAMSHIFT algorithm with Multidimensional Receptive Field Histograms was proposed in order to address some of the shortcomings of the original technique. The extension made the technique more resilient to conditions encountered within the context of piano playing and enabled a better segmentation when faced with harsh visual conditions.

Many of the approaches discussed within this paper are tailored for specific applications that rely on environmental assumptions and do not generally lend themselves to the more complex segmentation required in the context of musician's movements estimation. The results of the explored statistical methods support this reasoning and reveal the need for more robust techniques. The provision of quantitative performance measurements for both musicians and athletes is an important challenge in computer vision which has yet to be resolved.

Acknowledgements

The authors wish to acknowledge the financial support from the University of Ottawa and from the Natural Sciences and Engineering Research Council of Canada that made this work possible.

References

[1] L. Lucchese, and S.K Mitra, "Color Image Segmentation: A State-of-the-Art Survey," in *Proc. of the Indian National Science Academy (INSA-A)*, New Delhi, India, vol. 67 A, no. 2, pp. 207-221, Mar. 2001.

[2] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," in *Intl Journal of Computer Vision*, vol. 1, no. 4, pp. 321-331, 1987.

[3] C. Gu, and M.-C. Lee, "Semiautomatic Segmentation and Tracking of Semantic Video Objects," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 572-584, Sept. 1998.

[4] N. Peterfreund, "Robust Tracking of Position and Velocity with Kalman Snakes," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 6, pp. 564-569, June 1999.

[5] S. Sun, D.R. Haynor, and Y. Kim, "Semiautomatic Video Object Segmentation Using VSnakes," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 1, pp. 75-82, Jan. 2003.

[6] S. N. Krjukov, T. O. Semenkova, V. A. Pavlova, and B. I. Arnt, "Backpropagation Neural Network for Adaptive Color Image Segmentation," in *Proc. of SPIE, Applications of Artificial Neural Networks in Image Processing II*, vol. 3030, pp. 70-74, Feb. 1997.

[7] L. Xiong, D. Li, H. Hu, and G. Jin, "Segmenting the Color Image in a Simple Background by ANN Method," in *Proc. of SPIE, Symposium on Multispectral Image Processing*, vol. 3545, pp. 470-473, Oct. 1998.

[8] K. A. McCrae, D. W. Ruck, S. K. Rogers, and M. E. Oxley, "Color Image Segmentation," in *Proc. of SPIE, Applications of Artificial Neural Networks*, vol. 2243, pp. 306-315, Apr. 1994.

[9] A. Doulamis, N. Doulamis, K. Ntalianis, and S. Kollias, "An Efficient Fully Unsupervised Video Object Segmentation Scheme Using an Adaptive Neural-Network Classifier Architecture," in *IEEE Trans. Neural Networks*, vol. 14, no. 3, pp. 616-630, May 2003.

[10] S.-J. Lee, C.-S. Ouyang, and S.-H. Du, "A Neuro-Fuzzy Approach for Segmentation of Human Objects in Image Sequences," in *IEEE Trans. Systems, Man and Cybernetics*, vol. 33, no. 3, pp. 420-437, June 2003.

[11] D. Wang, "Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 539-546, Sept. 1998.

[12] T. N. Pappas, "An Adaptive Clustering Algorithm for Image Segmentation," in *IEEE Trans. on Signal Processing*, vol. 40, no. 4, pp. 901-914, Apr. 1992.

[13] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "Video Object Segmentation Using Bayes-Based Temporal Tracking and Trajectory-Based Region Merging," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, no. 6, pp. 782-795, June 2004.

[14] Y.-P. Tsai, C.-C. Lai, Y.-P. Hung, and Z.-C. Shih, "A Bayesian Approach to Video Object Segmentation via Merging 3-D Watershed Volumes," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 175-180, Jan. 2005.

[15] E.L. Andrade, J.C. Woods, E. Khan, and M. Ghanbari, "Region-Based Analysis and Retrieval for Tracking of Semantic Objects and Provision of Augmented Information in Interactive Sport Scenes," in *IEEE Trans. on Multimedia*, vol. 7, no. 6, pp. 1084-1096, Dec. 2005.

[16] D. Comaniciu, and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 750-755, June 1997.

[17] F. Bayoumi, M. Fouad, and S. Shaheen, "Based Skin Human Detection in Natural and Complex Scenes," in *Proc. of the IEEE Intl Midwest Symp. on Circuits and Systems*, vol. 2, pp. 568-571, Dec. 2003.

[18] W. Du, and H. Li, "Vision Based Gesture Recognition System with Single Camera," in *Proc. of the 5th International Conference on Signal Processing*, vol. 2, pp. 1351-1357, Aug. 2000.

[19] M.J. Jones, and J. M. Rehg, "Statistical Color Models with Application to Skin Detection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 274-280, June 1999.

[20] L. Sigal, S. Scarloff, and V. Athitsos, "Skin Color-Based Video Segmentation under Time-Varying Illumination," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 832-877, July 2004.

[21] J. Yang, W. Lu, and A. Waibel, "Skin-Color Modeling and Adaptation," in *Proc. of the 3rd Asian Conference on Computer Vision*, vol. 2, pp. 687-694, 1998.

[22] C. Stauffer, and W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 246-252, 1999.

[23] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-Time Tracking of the Human Body," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780-785, July 1997.

[24] C. Stauffer, and W.E.L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747-757, Aug. 2000.

[25] T. Horprasert, D. Hardwood, and L. S. Davis, "A Robust Background Subtraction and Shadow Detection," in *Proc. of the 4th Asian Conference on Computer Vision*, vol. 1, pp. 983-988, Jan. 2000.

[26] S. Atev, O. Masoud, and N. Papanikolopoulos, "Practical Mixtures of Gaussians with Brightness Monitoring," in *Proc. of the 7th IEEE Intl Conf. on Intelligent Transportation Systems*, pp. 423-428, Oct. 2004.

[27] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-Based Object Tracking," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564-577, May 2002.

[28] H. Tek, D. Comaniciu, and J. P. Williams, "Vessel Detection by Mean Shift Based Ray Propagation," in *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pp. 228-235, Dec. 2001.

[29] G. R. Bradski, "Computer Vision Face Tracking for Use in a Perceptual User Interface," in *Intel Technology Journal*, vol. 2, no. 2, 1998.

[30] M. J. Swain, and D. H. Ballard, "Indexing Via Color Histograms," in *Proc. of the 3rd Intl Conf. on Computer Vision*, pp. 390-393, 1990.

[31] B. Schiele, and J.L. Crowley, "Recognition without Correspondence using Multidimensional Receptive Field Histograms," in *Intl Journal of Computer Vision*, vol. 36, no. 1, pp. 31-52, 2000.

[32] F. Pelisson, D. Hall, O. Riff, and J.L. Crowley, "Brand Identification Using Gaussian Derivative Histograms," in *Proc. of the 3rd Intl Conference on Computer Vision Systems*, pp. 429-501, 2003.