

2005

Knowledge-based semantic image segmentation and global precedence effect

Fardin Akhlaghian Tab
University of Wollongong

Golshah Naghdy
University of Wollongong, golshah@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/infopapers>



Part of the [Physical Sciences and Mathematics Commons](#)

Recommended Citation

Tab, Fardin Akhlaghian and Naghdy, Golshah: Knowledge-based semantic image segmentation and global precedence effect 2005.
<https://ro.uow.edu.au/infopapers/2794>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Knowledge-based semantic image segmentation and global precedence effect

Abstract

This paper introduces a knowledge-based semantic image segmentation which extracts the "object(s)-of-interest" from the image. Image templates are the high-level knowledge in the system. The major contribution of this work is the use of the "Global Precedence Effect" (forest before trees) of the human visual system (HVS) in image analysis and understanding. The "object-of-interest" is searched for hierarchically through an irregular pyramid by an affine invariant comparison between the different region combinations and the template starting from lowest to the highest resolutions. The global/large size objects are found at lower resolutions with significantly lower computational complexity.

Disciplines

Physical Sciences and Mathematics

Publication Details

Akhlaghian, F. & Naghdy, G. (2005). Knowledge-based semantic image segmentation and global precedence effect. The 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT 2005) (pp. 237-244). United Kingdom: IEEE.

KNOWLEDGE-BASED SEMANTIC IMAGE SEGMENTATION AND GLOBAL PRECEDENCE EFFECT

Fardin Akhlaghian Tab and Golshah Naghdy

School of Electrical, Computer and Telecommunications Engineering
University of Wollongong, Wollongong, NSW 2522, Australia
Emails: {fat98, golshah}@uow.edu.au, Fax: +61 2 4227 3277

Keywords: Multiresolution segmentation, template, "object-of-interest" extraction, global precedence effect.

ABSTRACT

This paper introduces a knowledge-based semantic image segmentation which extracts the "object(s)-of-interest" from the image. Image templates are the high-level knowledge in the system. The major contribution of this work is the use of the "Global Precedence Effect" (forest before trees) of the human visual system (HVS) in image analysis and understanding. The "object-of-interest" is searched for hierarchically through an irregular pyramid by an affine invariant comparison between the different region combinations and the template starting from *lowest* to the highest resolutions. The global/large size objects are found at lower resolutions with significantly lower computational complexity.

1. INTRODUCTION

Visual information processing is becoming increasingly important with the advance of broadband networks, high power workstations, and advanced imaging tools such as digital cameras and scanners. Effective visual information management including saving, retrieval, distribution and presentation need an efficient processing method such as object-based processing. Therefore object-based processing has been used in many image and video processing applications such as image/video database and retrieval, coding, editing, interactivity, etc. The main challenge in implementing these processes is semantic segmentation and object extraction. Semantic segmentation and "object-of-interest" extraction in a general scene is not a trivial task and have received great attention in recent years [1-4].

The aim of segmentation is partitioning the image into semantic object(s)/region(s) for further processing. To date, perfect and Well implementation of this goal at this stage is far from reality [5], and the existing algorithms suffer from many limitations such as only coping with scenes with including special and limited objects [1, 6, 7]. Therefore "object-of-interest" extraction has been attracted great attention in recent years [2-4, 8].

There are many works on object extraction and recognition however, except the works which are designed for specific applications; very few consider a real segmentation and extraction stage. Some of the works in object recognition assume that the objects' shape are already extracted [9, 10], while others use a

simple segmentation algorithm by considering the object(s) in a very simple scene rather than a real image [11, 12]. These algorithms are more about recognizing and classifying the detected shapes rather than segmenting the object in a cluttered scene. This limitation originates from the gap between low level features and semantic concepts. However a generic comprehensive solution should include both object extraction and recognition.

Considering the gap between low-level features and semantic concepts, in any generic comprehensive semantic segmentation, high level knowledge is necessary. Therefore many semantic segmentation algorithms use some kind of template or high level features as knowledge base. The template is searched for through the segmented image. However, exhaustive search through the image has high computational complexity and there are not many effective search algorithms. Therefore effective search in the image for the "object-of-interest" is a topic in need of more research. In this work a comprehensive knowledge-based solution is presented which includes both low level segmentation and high level object extraction stages. The proposed algorithm facilitates the search in the image through defining a hierarchy of the examined objects/regions. This hierarchy originates from a feature in human visual system (HVS) called "global precedence effect" (GPE) [13]. The global and large size objects are examined/extracted before the local/small size objects. Therefore the analysis starts with the global information and local information are subsequently used to refine the decision making process. Multiresolution search through a pyramid is an effective solution for GPE implementation resulting in less complex search algorithm. In this paper a hierarchical search algorithm through a defined irregular pyramid is proposed.

For low level segmentation, an scalable pyramid segmentation is proposed which produces the same pattern at different resolutions [14]. The similarity of patterns increases the reliability of low resolution segmentation and search. However, due to insufficient information at very low resolution, an irregular pyramid is introduced which allows search through very low resolution and reduces the computational complexity significantly.

This paper is organized as follows. Section 2 presents a short review of the related works. In Section 3, the scalable pyramid segmentation is briefly explained. The single resolution search for the "object-of-interest" is described in Section 4. The computational complexity is also discussed in this Section. The Hierarchical search through irregular pyramid is pro-

posed in Section 5. Some experimental result and discussion are presented in Section 6, and finally conclusions are drawn in Section 7.

2. RELATED WORKS

We separate the works in semantic segmentation literature into two categories. In the first group, the “object-of-interest” is searched for by considering the low level information in the segmented image. Moreover, some high level knowledge about the objects’ characteristics such as object’s model and qualitative and quantitative relationships are often employed.

In [2] the image is segmented into different regions and the curve of the “object-of-interest” is transformed into the affine invariant and is searched for through the image for the best match. The transform parameters are changed until the best match in the image is found. The algorithm requires exhaustive search in the parameters and image spaces resulting in high computational complexity.

Xu et al. [4] segment the image into homogenous regions and all possible combinations of various regions are affine invariant matched with the template. A group of regions with Hausdorff distance less than a threshold represent a possible “object-of-interest”. Due to a great number of possibilities for combined regions, the search is computationally very complex. To lighten the computational burden a stack of different segmentation maps is proposed. The two most similar regions are merged to create the next segmentation map in the stack. The merging continues until a segmentation map with two regions is reached. The search starts from the segmentation map with two regions and continues through other segmentation maps until the “object-of-interest” is found. The deficiency of this algorithm is that the number of segmentation maps depends on the number of regions and can be very large and two consecutive segmentations maps in the stack are very similar. Finally the merged regions are based on the statistical similarity criterion while better criteria such as semantic criterion can be used.

In the second group of algorithms, using the low level features of the image such as colour, texture, edges, etc the image regions are extracted, refined and combined to establish their correspondence to a higher level image descriptions. For example, regions belonging to same class such as grass or sky are mixed together. These algorithms don’t guarantee that the final regions are all semantic regions representing meaningful objects or regions. In many works the detected objects are often rigid and simple such as sky, water, etc., belonging to a homogenous region. Therefore many of these algorithms are about natural image segmentation [1, 7]. An example of the application of these algorithms is remote sensing [8]. These algorithms often cannot extract a complex object such as human, car, etc., in a real image.

3. SCALABLE PYRAMID SEGMENTATION

The proposed spatial segmentation fits multiresolution Markov random field (MMRF) image segmentation with the spatial scalability [14]. Images at different resolutions are segmented with spatial scalability as a constraint which keeps the 4 : 1

pixels down sampling relation between different resolution segmentations. Therefore the produced segmentation maps are similar at different resolutions.

To extend the single level Markov random field (MRF) based segmentation [15] to a multiresolution scalable segmentation (SSeg) algorithm, it should be noted that the corresponding pixels at different resolutions have the same segmentation classification. Therefore the classification of these pixels changes together and they are processed together in a multidimensional space. Consequently, objective function of regular single level Bayesian segmentation [15] is extended to a multidimensional space by the following equation:

$$E(X) = \sum_{\{s\}} \{ \|Y(\{s\}) - \mu^{X(\{s\})}(\{s\})\|^2 + \sum_{\{\tau\} \in \Theta(\{s\})} V_c(\{s\}, \{\tau\}) \} \quad (1)$$

In this expression, s is a pixel of the pyramid decomposition and $\{s\}$ is the set including s and its corresponding pixels¹ in other resolutions. Y , X and μ are intensity, intensity segmentation and intensity average functions respectively. V_c is the clique function defined on two neighbouring sets of corresponding pixels. In regular single level segmentation, cliques are defined over two adjacent pixels s and τ by the following formula:

$$V_c(s, \tau) = \begin{cases} -\beta & \text{if } X(s) = X(\tau) \\ +\beta & \text{if } X(s) \neq X(\tau) \end{cases} \quad (2)$$

In the proposed scalable multiresolution segmentation algorithm, all the corresponding pixels of $\{s\}$ at different resolutions are examined with their neighbouring pixels through the following equation:

$$V_c(\{s\}, \{\tau\}) = \binom{1}{N} \sum_{k=M}^{M+N-1} (-1)^{L_k} \beta, \quad (3)$$

$$L_k = \begin{cases} 1 & \text{if } X(s_k) = X(r_k) \\ 0 & \text{if } X(s_k) \neq X(r_k) \end{cases} \quad \& \quad s_k \in \{s\}, r_k \in \{\tau\}$$

In (3), M is the lowest resolution in the pixels of $\{s\}$ and N is the number of different resolutions of pixels in $\{s\}$. The first summation in (1) is over all pixel set of corresponding pixels at different resolutions and the second one is over all the cliques including the set $\{s\}$.

For the optimization of MMRF modelling, the Iterated Condition Mode (ICM) algorithm matched to the scalable multiresolution segmentation is used. The energy function of equation (1) is optimized iteratively from lower resolution to higher resolutions. More explanation about this spatial segmentation algorithm can be found in [14].

¹Corresponding pixels are related to each other by downsampling pixels of higher to lower resolutions. Therefore each pixel has corresponding pixels at higher resolutions, but not necessarily at lower resolution.

4. OBJECT OF INTEREST EXTRACTION AT SINGLE RESOLUTION

“Object-of-interest” extraction is often based on the minimization of a suitable distance between a reference such as a template and a grouping of regions in the test image. Template matching is an approach to recognize the “object-of-interest” in the digital images. In a real scenario, the “object-of-interest” are searched in a segmented image². Each possible region combination is examined by the region matching algorithm. Therefore, due to the huge number of possible region combinations, a simple shape matching algorithm is preferred. In this work a region-based shape matching is proposed which is a combination and modification of the two approaches presented in [16, 17].

The comparison should be scale, rotational and translation invariant therefore the first stage of comparison is variation compensation. At first the shape rotation is compensated. The idea is to find the major axes of the two shapes. The major axis is a straight line which connects the two farthest pixels on the shape’s border. The angle between the two axes determines the rotational angle factor. The template is rotated so that its major axes lie in the same direction as the candidate region’s major axes. The ratio between the two major axes’ lengths determines the scale normalization factor. The shape’s size is normalized by the scale factor. Using similar scaling approach, the shape in a major and minor directions is scaled so that both shapes have the same bounding box which is the smallest rectangle containing the shape. Then the bounding box areas of the two shapes are translated to the origin. Finally, Hausdorff distance measures the distance between two sets of binary images’ pixels [17]. The similarity value is computed and based on a user defined threshold, is accepted or rejected.

Generally, if it be supposed that the image is segmented to N fully connected regions, the maximum number of possible region combinations is:

$$\sum_{k=1}^N \binom{N}{k} = 2^N - 1 \quad (4)$$

This is a very big number for normal values of segmentation regions such as $N = 50$. Of course this is the worse case which assumes that all regions are connected together and all combinations are examined. Practically neighbouring is a local feature and the number of possible combinations is much less than $2^N - 1$. However, the experimental results in Section 6 show that computational complexity is so high that it practically renders the algorithm useless for real applications.

5. HIERARCHICAL SEARCH OF OBJECT OF INTEREST

Inspired by a well known feature in the human visual system called “global precedence effect” (forest before trees) where the processing pathway for outline (low frequency) is faster than detail (high frequency) [13, 18], a hierarchical search is

²Any segmentation algorithm can be used, but segmentation with less region with the successful separation of interested objects/regions from background is more desired.

proposed. In a simple way, low resolution image where the outline of the “object-of-interest” is defined is given higher priority in the search process and if the search fails higher resolutions are searched until the search process is exhausted. For template search we need to perform segmentation at different levels. This is done by the multiresolution scalable segmentation algorithm proposed in [14]. The scalability of the proposed segmentation is a valuable feature at this stage because it maintains the similar segmentation patterns at different resolutions. This increases the accuracy and reliability of the search at the lower resolutions. Furthermore, the perfect relation of parent and children between regions at different resolutions will detect the extracted object at other resolutions. However, due to insufficient information, the search at low or very low resolution such as 4×4 pixels is not accurate or useful. To remove this problem a stack is proposed which keeps the image size and gives different priorities to different regions groupings. The defined stack is a combination of *full size* image segmentation maps which correspond to the segmentation at different resolutions of the pyramid. This stack is called an irregular pyramid. The elements of the stack or irregular pyramid are built hierarchically from fine to lowest resolution. At each resolution, the hierarchical segmentation (HSeg) is obtained by considering three other segmentations; 1) the corresponding regular pyramid segmentation at the same resolution 2) Regular pyramid segmentation of the neighboring finer resolution and 3) Irregular pyramid segmentation of the neighboring finer resolution. Figure 1 shows this relationship.

The base of the stack (irregular pyramid) is identical to the finest resolution at the top of the regular scalable pyramid segmentation. On going down through pyramid toward lower resolutions, small objects/regions are deleted, and the number of existing regions decreases. Similarly, these regions should be deleted from the corresponding irregular pyramid (hierarchical) segmentation. The size reduction during the pyramid decomposition deletes small regions physically. However, in the irregular pyramid segmentation, the size is kept the same and the physical deletion of regions doesn’t occur. Therefore the regions are deleted logically: the deleted regions are merged with the neighbouring regions by a criterion such as similarity and the existence of salient edges between regions. Practically, the regular pyramid guides the hierarchical segmentation to delete m regions hierarchically in n steps, where n is the number of levels in the pyramid-based decomposition and $m \gg n$. As we proceed towards the lower resolutions of the regular pyramid, the correspondings smaller regions in the irregular pyramid are (logically) deleted and global regions related to low spatial frequency with larger size objects remain. Finally, at the lowest level of the pyramid there is the hierarchical full size segmentation with only two regions at the top of the stack.

The search starts through the hierarchical segmentation patterns at the top of the stack. If the “object-of-interest” is not found at the current resolution, the hierarchical segmentation patterns corresponding to the next higher scale will be popped from the stack and it will be searched for the “object-of-interest”. The search will continue through higher scale hierarchical segmentation image until the “object-of-interest” is found. The lower resolution region combinations have corresponding regions at the hierarchical segmentation of higher

resolutions. Therefore they need not to be tested at the higher resolutions. Only newly emerging region combinations at the higher resolutions are tested.

It is clear that the proposed hierarchal object search detects global and large size objects much faster than the regular single resolution search. However, the computational savings for the detection of the local and small size objects is minimal. Nevertheless this priority search for the detection of "object-of-interest" is more efficient and is consistent with the human visual system. In many applications such as object-based image retrieval, the "object-of-interest" is the global and main subject of the image. If the initial scrutiny of the global information in the image, does not detect the "object-of-interest", the processing can proceed to the next step which include the analysis of local or finer resolutions until the search is exhausted. In these cases the proposed search significantly reduces the computational complexity and facilitates a more effective "object-of-interest" extraction.

6. EXPERIMENTAL RESULTS AND DISCUSSION

To show the full benefit of the use of the "global precedence effect" and the advantages of the hierarchical object extraction, this section presents the simulation results on some real images including "head and shoulder", car, etc. The shape matching algorithm described in section 4 is utilized to measure the similarity between the shape template and the candidate regions. Because each example has many images at different resolutions, they are shown by equal small size figures. The results are discussed and the advantageous/disadvantageous of the proposed multiresolution segmentation and hierarchical search are illustrated.

As an example, a relatively simple image of the first frame of CIF size sequence of foreman image sequence is chosen. The image is CIF size with YUV color format where Y is in full resolution and U and V are in half resolution. In many of the video object tracking algorithms, a semi automatic processes such as user intervention and fine tuning is used to, detect the "object(s)-of-interest" at the first frame. [19–21]. The proposed object detection algorithm, however, can be used for automatic extraction of the "object(s)-of-interest" from the first frame of image sequences. The original image is shown in image 2 (a). The decomposed pyramid images are segmented by the proposed scalable segmentation. The scalable segmentation and hierarchical segmentation are also shown in Figures 2(b) to (r). The 9×11 is the lowest resolution that the "object-of-interest" is effectively separated from the image background area. Therefore the "object-of-interest" is searched from low to high resolutions in 2×2 , 3×3 , 5×6 , and 9×11 resolutions respectively. The maximum number of candidate regions will be:

$$\sum_{k=1}^{21} \binom{21}{k} + \sum_{k=1}^{16} \binom{16}{k} + \sum_{k=1}^7 \binom{7}{k} + \sum_{k=1}^3 \binom{3}{k} = 1114244 \quad (5)$$

The number of regions and the region combinations are shown in Table 1. The real number of region combinations in the four lowest resolution include 2×2 , 3×3 , 5×6 and

9×11 is equal to $3298 + 732 + 43 + 7 = 4090$. Therefore the real number of search is less than 4090 regions and is greatly less than 1114244 regions. As the Table 1 indicates, from the resolution 18×22 toward higher resolutions the number of region combinations increases so much that practically it is impossible to search for the "object-of-interest" over these resolutions. In particular, at the highest resolution the number of region combinations is so high that the search is practically impossible. The efficiency of the pyramidal template search compared to single resolution template search is $(1 - 4090/(1.37 \times 10^8)) \approx \%99.99$ which is very close to $\%100$. Regular single resolution produces more regions than regular multiresolution segmentation and the proposed scalable pyramid segmentation algorithms [14]. This increase in the number of regions increases the computational complexity of the search algorithm.

The extracted "object-of-interest", its template and the regions matching with the template are shown in Figures 3 (a) to (e). The Hausdorff distance between the object's template and the extracted object is 7.4. An example of a rejected region, a region and its match with the template model is also shown in Figures 3 (f) to (h). The Hausdorff distance of this tested object and template is 30.65. A threshold that pass 7.4 and reject the other regions Hausdorff distance such as 30.65 should be entered into the algorithm.

In the next example the detection of a small size object is considered. The original image is seen in Figure 4 (a). The grey level image is in SIF size and the "object-of-interest" is the ball, which is a small size object. The image is decomposed to 10 different scales by the wavelet decomposition. The image pyramid is then segmented by the scalable segmentation. The scalable and its corresponding hierarchical image segmentation at the different resolutions can be seen in Figure 4 (b) to (p). Due to the small size of the "object-of-interest", it is not detected before the 5th level of pyramid decomposition. Therefore the resolutions 1×2 , 2×3 , 4×6 , 8×11 are searched, and finally the "object-of-interest" is found at the 15×22 resolution. This hierarchal search, from global to local information, is quite consistent with the "global precedence effect". The template, the found region and their match are shown in Figure 5 (a) to (d). The Hausdorff distance of the match is 4.62. Table 2 shows the number of regions and their combinations. $3 + 15 + 78 = 96$ region combinations are searched at the three resolutions lower than 15×22 and the 1058 combinations at this resolution which the object is found. Therefore in total $96 + 1058 = 1152$ region candidates are searched to find the "object-of-interest". From this number $96/1152 * 100 = 8.3\%$ of regions are searched at lower resolutions.

This example shows that the detection of small size objects is done at higher resolutions of the pyramid with more complexity than the large size objects at lower resolutions. But this is an acceptable property consistent with the "global precedence effect" of the human visual system.

7. CONCLUSION

In this paper a hierarchical region-based image object extraction and recognition/classification algorithm is proposed. Simulating the "global precedence effect" of human visual system

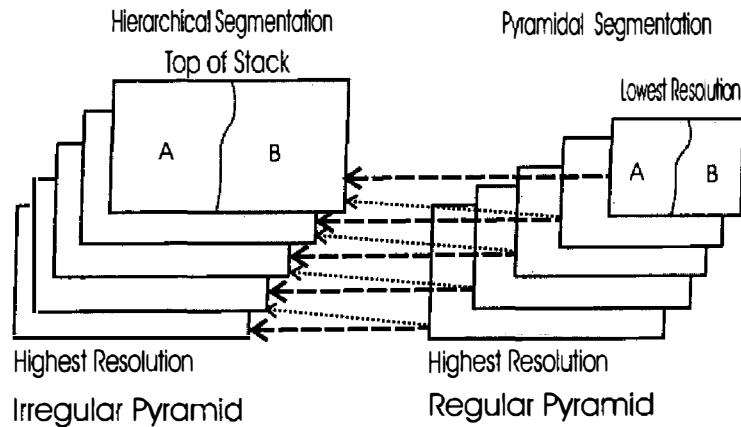


Figure 1: The hierarchical stack or irregular pyramid segmentation corresponding to the pyramid segmentation. Only two regions are on segmentation the top of the stack.

Table 1: Number of regions and combinations at resolutions of foreman segmentation.

Resolution	288 × 352	144 × 176	72 × 88	36 × 44	18 × 22	9 × 11	5 × 6	3 × 3	2 × 2
Number of Regions	89	85	81	69	61	21	16	7	3
Combinations number	8.37×10^7	2.01×10^7	9.5×10^6	5.94×10^5	174749	3298	732	43	7

Table 2: Number of regions and combinations in the Table Tennis segmentation.

Resolution	240 × 352	120 × 176	60 × 88	30 × 44	15 × 22	8 × 11	4 × 6	2 × 3
Number of Regions	79	62	45	27	18	11	4	2
Combinations	8.93×10^8	1.83×10^6	2.12×10^5	7019	1058	77	16	3

results in a hierarchy of objects and significantly decreases the number of tested candidate regions and computational complexity. The proposed hierarchical segmentation patterns organized in an irregular pyramid allows detecting the main global object at the lower resolutions with less computational complexity while small objects are detected at higher resolutions with higher computational complexity. The proposed recognition needs the template of the “object-of-interest” which is a high level knowledge about the “object-of-interest”. Template design, especially deformable templates, and reducing the complexity of matching algorithms needs more research. The suitable threshold for decision about accepting or rejecting a region as the “object-of-interest” is tuned by the user and its automatic setting needs further research. The proposed algorithm is a significant step towards object extraction in real images.

REFERENCES

- [1] L. F. C. K. S. Newsam, S. Bhagavathy and B. S. Manjunath, “Object based representations of spatial images,” *Acta Astronautica*, vol. 48, p. 567577, 2001.
- [2] F. Marques, M. Pardas, and R. Morros, “Object matching based on partition information,” in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 2, pp. II-829–II-832.
- [3] J. Fan, D. K. Y. Yau, A. K. Elmagarmid, and W. G. Aref, “Automatic image segmentation by integrating color-edge extraction and seeded region growing,” *Image Processing, IEEE Transactions on*, vol. 10, pp. 1454–1466, 2001.
- [4] Y. Xu, E. Saber, and A. M. Tekalp, “Dynamic learning from multiple examples for semantic object segmentation and searches,” *Computer Vision and Image Understanding*, vol. 95, pp. 334–353, Sept. 2004.
- [5] R. Alferez and Y. F. Wang, “Image database indexing using a combination of invariant shape and color descriptions,” in *Proceedings of the Second International Conference on Information Fusion. FUSION '99. Int. Soc. Inf. Fusion. Part vol.2, 1999*, pp.688-95 vol.2. Mountain View, CA, USA.
- [6] S. Lefevre, L. Mercier, V. Tiberghien, and N. Vincent, “Multiresolution color image segmentation applied to background extraction in outdoor images,” in *Proceedings CGIV'2002. First European Conference on Colour in Graphics, Imaging, and Vision. Soc. Imaging Sci. & Technol. 2002*, pp.363-7. Springfield, VA, USA.
- [7] J. Boldys, “Bayesian supervised segmentation of objects in natural images using low-level information,” in *Image and Signal*

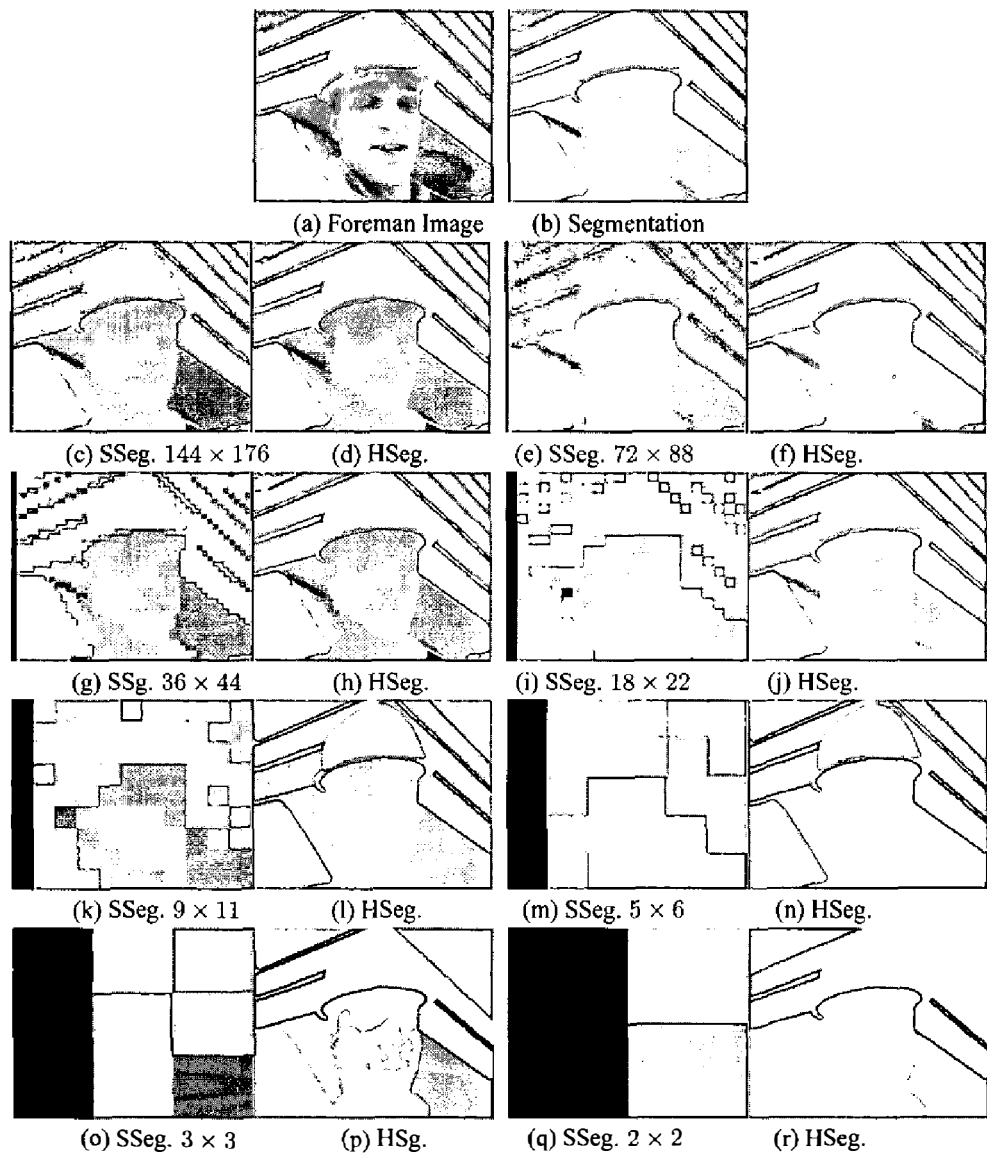


Figure 2: Foreman original image and its scalable segmentation (SSeg) and hierarchical segmentation (HSeg) at different resolutions. The hierarchical segmentation images are just after the scalable segmentation at any resolution; (a) The original image at 288×352 resolution; (b) 288×352 segmentation; (c) 144×176 segmentation; (e) 72×88 segmentation; (g) 36×44 segmentation; (i) 18×22 segmentation; (k) 9×11 segmentation; (m) 5×6 segmentation; (o) 3×3 segmentation; (q) 2×2 segmentation;

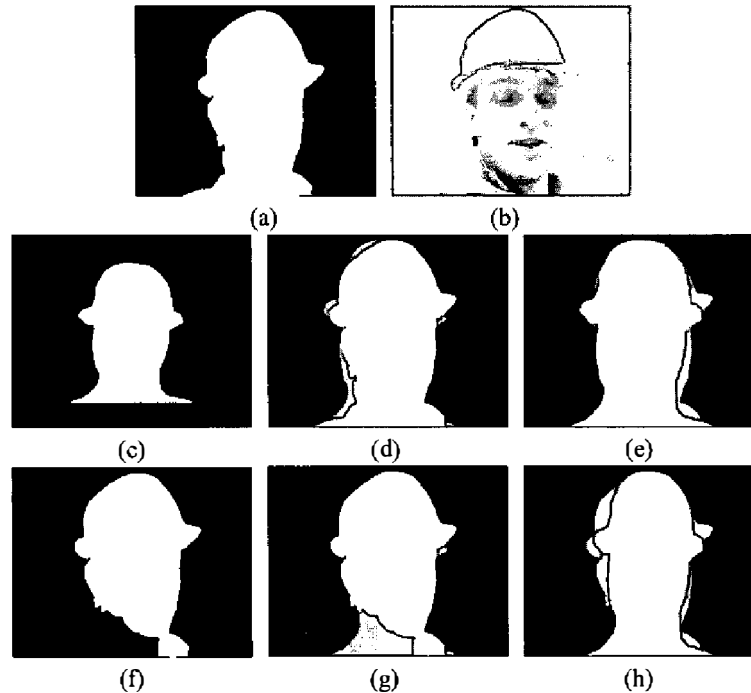


Figure 3: (a) The extracted Foreman head and shoulder's shape; (b) The extracted Foreman head and shoulder's texture;(c) Template; (d) match between the template and the region, where the candidate region is drawn over the template; (e) template is over candidate region; (f) a (rejected) candidate region (g) match between the template and the candidate region, where the candidate region is drawn over the template; (h) template is over the regions;

- Processing and Analysis, 2003. ISPA 2003. Proceedings of the 3rd International Symposium on*, vol. 2, pp. 1054–1059.
- [8] A. Darwish, K. Leukert, and W. Reinhardt, "Image segmentation for the purpose of object-based classification," in *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, vol. 3, pp. 2039–2041.
- [9] F. A. Cheikh, A. Quddus, and M. Gabbouj, "Multi-level shape recognition based on wavelet-transform modulus maxima," in *Image Analysis and Interpretation, 2000. Proceedings. 4th IEEE Southwest Symposium*, pp. 8–12.
- [10] P. Brault and H. Mounier, "Automated transformation invariant, shape recognition through wavelet multiresolution," in *SPIE-Int. Soc. Opt. Eng. Proceedings of Spie - the International Society for Optical Engineering*, vol.4478, 2001, pp.434-43. USA.
- [11] M. Ashrafuzzaman, M. M. Rahman, and M. M. A. Hashem, "Automatic acquisition of object using eigen appearance," in *SICE 2002. Proceedings of the 41st SICE Annual Conference (Cat. No.02TH8648). Soc. Instrument & Control Eng. (SICE). Part vol.2, 2002*, pp. 1325–6.
- [12] C. Thies, V. H. Metzler, and T. Aach, "Content-based image analysis: object extraction by data-mining on hierarchically decomposed medical images," in *SPIE-Int. Soc. Opt. Eng. Proceedings of Spie - the International Society for Optical Engineering*, vol.5032, 2003, pp.579-89. USA.
- [13] D. Navon, "Forest before trees: The precedence of global features in visual perception," *Cognitive Psychology*, vol. 9, p. 353383, 1977.
- [14] F. A. Tab, G. Naghdy, and A. Mertins, "Multi resolution image segmentation with border smoothness for scalable object-based wavelet coding," in *Proc. 7th international conference on Digital Image Computing - Techniques and Applications (DICTA)*, Sydney, Australia, 2003, pp. 977–986.
- [15] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Transactions of Image Processing*, vol. 40, no. 4, pp. 901–914, Apr. 1992.
- [16] A. Sajjanhar and G. Lu, "Indexing 2d non-occluded shapes for similarity retrieval," in *SPIE-Int. Soc. Opt. Eng. Proceedings of SPIE - the International Society for Optical Engineering*, vol. 3164, 1997, pp.188-97.
- [17] Y. Xu, E. Saber, and A. M. Tekalp, "Object segmentation and labeling by learning from examples," *Image Processing, IEEE Transactions on*, vol. 12, pp. 627–638, 2003.
- [18] S. Han, E. W. Yund, and D. L. Woods, "An erp study of the global precedence effect: the role of spatial frequency," *Clinical Neurophysiology*, vol. 114, pp. 1850–65, 2003.
- [19] B. Marcotegui, F. Zanoguera, P. Correia, R. Rosa, F. Marques, R. Mech, and M. Wollborn, "A video object generation tool allowing friendly user interaction," in *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348). IEEE. Part vol.2, 1999*, pp.391-5 vol.2. Piscataway, NJ, USA.
- [20] D. K. Park, H. S. Yoon, and C. S. Won, "Multiresolution-based watersheds for efficient image segmentation," *Pattern Recognition Letters*, vol. 24, no. 1-3, pp. 473–488, 2003.
- [21] F. A. Tab, G. Naghdy, and A. Mertins, "Well defined video object extraction suitable for scalable wavelet based object coding," in *Signal Processing and Communications. 2004. SPCOM '04. 2004 International Conference on*, pp. 204–208.

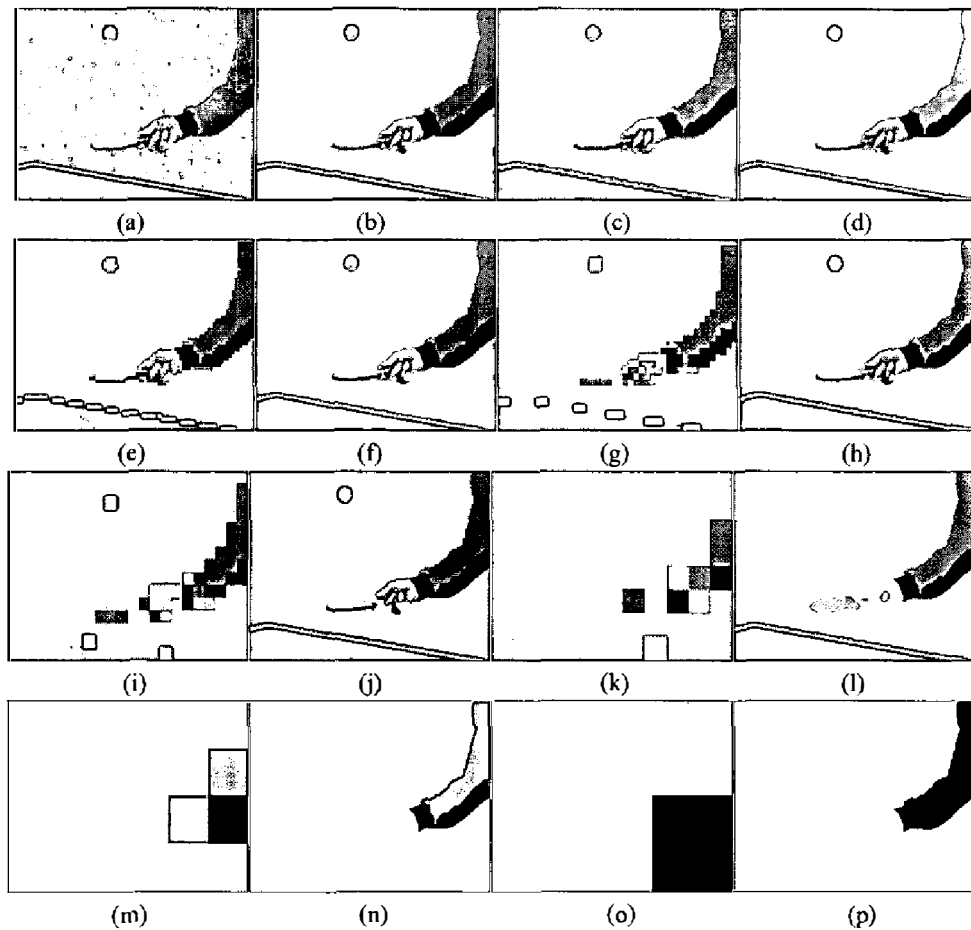


Figure 4: Table Tennis original image with its scalable segmentation (SSeg) and hierarchical segmentation (HSeg) at different resolutions. The HSeg images are just after the SSeg at each resolution: (a) the original image at 240×352 resolution; (b) 240×352 SSeg; (c) 120×176 SSeg; (d) HSeg corresponding to 120×176 ; (e) 60×88 SSeg; (f) HSeg corresponding to 60×88 SSeg; (g) 30×44 SSeg; (h) HSeg corresponding to 30×44 ; (i) 15×22 SSeg; (j) HSeg corresponding to 15×22 ; (k) 8×11 SSeg; (l) HSeg corresponding to 8×11 ; (m) 4×6 SSeg; (n) HSeg corresponding to 4×6 ; (o) 2×3 SSeg; (p) HSeg corresponding to 2×3 .



Figure 5: (a) The ball template; (b) The extracted ball shape at 15×22 resolution; (c) match between the template and the extracted ball, where the candidate region is drawn over the template; (d) template is over candidate region.