IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

EXTRACTION OF STANDING HUMAN BODIES FROM IMAGES WITH MULTI LEVEL SEGMENTATION AND SPLINE REGRESSION

R Swathi*, K Ramesh
*Electronics and Communication Engineering, JNTUACEA, India

ABSTRACT

The ability to extract and detection of human activities by computer vision is very important with many potential applications. Extraction of human bodies from images from respective digital image has accomplished consideration in recent times and extensive variety of research is carried on to meet the sought result. In this paper proposed a novel strategy to extract human bodies from images where the scene density, the highly dimensional pose space, and various human appearances are handled in better way compared to conventional state of art methods. The proposed approach is ordered into five distinct strides (i) face detection, (ii) Multiple level segmentation, (iii) skin detection, (iv) upper body segmentation and (v) lower body segmentation respectively. In this paper we propose another method spline regression to extraction of human bodies from single images. Finally the simulation results have achieved better performance and high proficiency over traditional state of art methods.

KEYWORDS: Face detection, Skin detection, upper body segmentation, Multiple level segmentation, lower body segmentation, spline regression.

INTRODUCTION

Extracting human bodies in cluttered images is difficult due to several factors, including shading, occlusions, image noise, background clutter, and the unrestricted positions due to in and out of the image plane rotations. To know about the human body region can uses various tasks, such as recognition of actions from still images ([2]-[5]), determination of the layout of human ([6]-[8]), and sign language recognition ([9]). The silhouette extraction is a challenging task, especially when we are considering intricate cases.

In this paper, we propose an analytical approach for human body segmentation in still images. We classified the problem into three problems: there is a direct correlation among them, Face detection, upper body extraction, and lower body extraction. In that face detection provides crucial indication about the presence of the humans in still images. It greatly decreases the search space for the upper body, and provides information about skin color. Face dimensions also hint in determination the dimensions of the rest of the body.
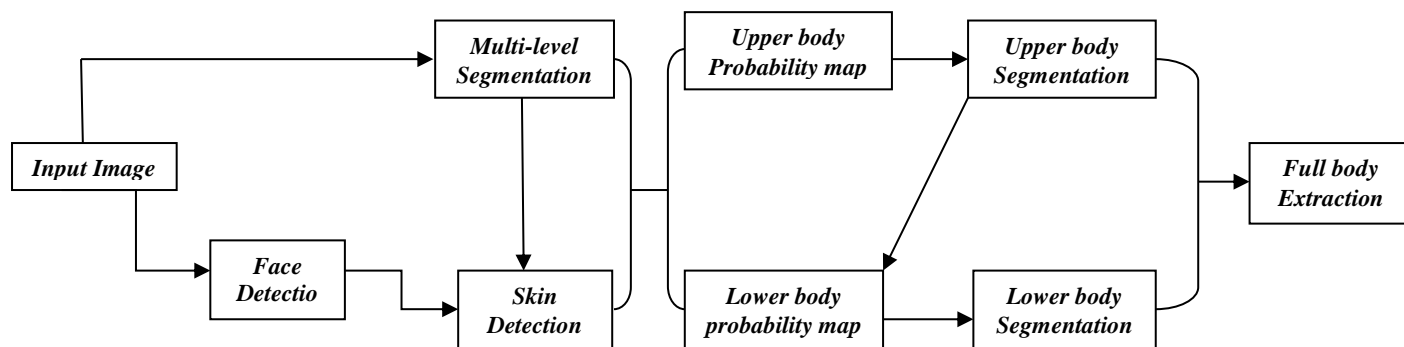

Fig. 1. Overview of proposed methodology.

Upper body extraction gives extra data about the position of the hands. Skin identification data manages look for the upper region, which in turns drives the scan for the lower body.

The steps followed in proposed method:

1) Proposed a new framework for automatic of segmentation human bodies in single images.
2) Combined information gathered from different levels of segmentation, which allows strong and efficient computations upon group of pixels that are ken correlated.
3) A soft anthropometric constraint allows the whole process and uncovers body regions.
4) Taken some assumptions that sleeves are of similar color of torso region, and the lower part of the pants is similar to the upper part of the pants, structured searching and extraction algorithm based on the premise that colors in body regions appear strongly in foreground and weakly in background region.

In this paper we extend the proposed algorithm for extraction human body in images using spline regression.

RELATED WORK

Human body segmentation approaches classify into four categories, Interactive methods, top-down approaches, bottom-up approaches, and Hybrid methodologies. In interactive methods, it expects user input in order to differentiate the background and foreground. Interactive methods give very accurate results in conflict cases. But these are not employing object specific knowledge, they often require user to guide their process and are not suitable for many real world problems. It's differed from automatic and often task specific.

The top down approaches are based on prior knowledge (or) anterior knowledge and utilize the image substance to assist refine an underlying model. The top down methodologies have been proposed as answers for the issue of fragmenting human bodies from static images. The principle normal for these methodologies is that they require abnormal state learning about the foreground, which in the case of people is their pose. One strategy for posture estimation and object recognition is the pictorial structures (PS) model and its varieties. However they depend on high level models that may come up short in complex cases, limiting the accomplishment of the final products. High level models are time consuming and usually these methods are computationally costly.

Bottom up methodologies utilize low level components, for example, pixels or super pixels and attempt to gathering them into semantic substances of higher levels. In various levels of segmentation chain of command are utilized, which are computationally costly. In this approach join prompts from different levels of segmentation. Posture estimation can be considered as a more elevated issue contrasted and body segmentation and numerous want to utilize a bottom up approach to deal with encourage pose recognition and body part estimation.

Hybrid philosophies influence the bottom up and top down methodologies. Perceptual groupings from a bottom up approach as a rule give a decent establishment to adapt with the high number of postures. Shape templates of body parts connected on image segments are a method for understanding the segmentation bewilder. Keeping in mind the end goal to ease the requirement for various shape templates, disintegrate the issue into upper and lower body estimation, also utilize fitting of an exact torso model. In two deformable models are intended to segment the human body on two scale super pixels, also a coarse to fine technique is utilized. *Li et al.* develop these ideas by consolidating a kinematic model and information driven graph cut. Rauschert and Collines exhibit Bayesian structure for together assessing explaining body pose, and the pixel level segmentation of every body part is proposed. The after effects of the strategy is extremely encouraging, however to handle the many-sided quality of the issue, not withstanding to standing positions, the proposed display should upgraded over large number of parameters.

PROPOSED METHOD
FACE DETECTION

Detection of the face region in proposed method is performed using OpenCV's implementation of the Viola-Jones algorithm that provides high speed and performance. Due to skin probability in proposed method is learned from the face region adaptively, so proposed method preferred an algorithm that is based on structural features of the face. The Viola-Jones face detector is leads to false positive detections that can lead to unnecessary activations of algorithm and falsie skin detections. To reduce to a fine state of the results of the method, proposed a skin detection algorithm based on color constancy and a multi layer perceptron neural network trained on still images collected under various illumination conditions both indoor and outdoor, the face detection is based on facial features detection and the act of localizing using image segmentation, graph based verification of the facial structure, and low level image processing techniques.

In the proposed method primarily, the pixels that similar to skin are detected using the method to image chromatic adaptation using neural networks. And then, the elliptical regions of the detected face in the images established by the Viola Jones method. In this paper considered that the average skin probability of the pixels X of face region Fr_i, for each person i, compared with threshold $T_{GlobalSkin}$ (taken empirically to 0.7 in proposed method). If this passes the global skin threshold that means greater than $T_{GlobalSkin}$, then it is evaluated by face detector. If the facial regions are detected, then face region is considered as a true positive detection.

Using fundamental unit, that means palm length the locations and sizes of human body parts are estimated, based on anthropometric constraints. Palm length referred that the major axis of the ellipse is almost same size as the distance from the base of the palm to the tip of the middle finger. In proposed method the anthropometric model is adaptive for each person and constant to scale.

MULTILEVEL SEGMENTATION

Depending exclusively on free pixels for confounded inference prompts to propagation of blunders to large amount of image processing in complex situations. There are a few distinctive sources of noise, such as compression, the digital sensors that captured the image or even the unpredictability of the image itself and their impact is more extreme at pixel level. A typical practice to ease the noise abiding at the pixel level is the utilization of filters and algorithms that extract aggregate data from pixels. Additionally, gatherings of pixels express higher semantics. Little gatherings safeguard and vast groups tend to catch shape and more theoretical structures better. At long last, the computations in view of superpixels are more proficient and encourage more adaptable algorithms.

In this proposed scheme, propose utilizing a image segmentation technique, with a specific end goal to process pixels in more important groups. Be that as it may, there are various image segmentation algorithms, what's more, the determination of a proper one depend on the accompanying criteria. To begin with, we require the algorithm to have the capacity to save solid edges in the image, since they are decent sign of boundaries between semantically extraordinary regions. Second, another alluring characteristic is the creation of segments with generally uniform sizes. More vitally, proposed a utilizing various levels of segmentation, keeping in mind end goal to lighten the requirement for selecting proper number for regions to be made and consolidate information exuding from various perceptual groupings of pixels. Despite the fact that proposed system can acknowledge any of segmentation levels, proposed system find that two segmentation levels of 100 and 200 segments give precise results.
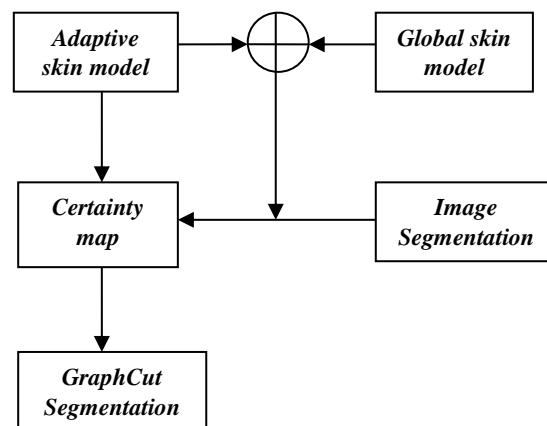
SKIN DETECTION

Fig. 2. Skin detection algorithm

Skin tone varies because of illumination and ethnicity, the fact that limbs often do not contain enough contextual information to discriminate, and skin like regions are the most prominent obstacles to detecting skin regions in images. In this paper, proposed a combining an appearance model created for each face with global detection. It provides strong distinction between skin and skin like pixels. These segmentation cues are used to create regions of ambiguous. Regions of certainty and ambiguous consist a map that follows the GrabCut algorithm, which provides outputs for final skin regions. Using anthropometric constraints and body connectivity are eliminating false positives.

In proposed method each face region Fr_i, is used to construct an adaptive color model for each person's skin color. In this study, proposed using r, g, s, I, Cr and a channels. In detail, r = R/(R+G+B), g = G/(R+G+B), and s = (R+G+B)/3; here r and g are normalized versions of R and G channels, and s is used instead of b to get channel independence. Image pixel's probability of being skin pixel is calculated sunder ling for each channel according to a normal probability distribution. The skin probability for each pixel X is as follows:

$$P_{Skin_i}(X) = \prod_{j=1}^{6} \mathcal{N}(X, \mu_{ij}, \sigma_{ij}) \qquad (1)$$

The global model and the adaptive model are combined through weighted averaging (a weight of 0.25 for global, and 0.75 for adaptive model). In order to characterize a region as a probable foreground or background, its mean probability of the combined probability must above a certain threshold (empirically set to 0.2 and 0.3, respectively).

UPPER BODY SEGMENTATION

The early and most important step in proposed methodology is face detection, which navigates the rest of the process. The extracted information in this step is very important and significant. In this first, the color of the skin in a face can be used to identify or match the rest of the visible skin parts of person's body. Second, the location of face gives the strong cue up about the rough location of the torso. In this case assumption about torso is below the face region, but without strong assumptions about out and in of plane rotations. Third, the size of the face region leads to the estimation of the size of other body parts using anthropometric constraints.

After elliptical region of the region is known, then proceed for the foreground probability estimation. To better employ the existing spatial and color relations of the image pixels, perform multiple levels over segmentation. In respect of clothes, the size of face's ellipse guides the building of the rectangular masks for the foreground using anthropometric constraints. In this proposed method basic assumption is that effective foreground mask should consist regions that appear mainly inside the mask and not background (outside) that means identify the "islands of saliency" above mentioned sense. The different levels of segmentation give different perceptual pixel groupings, and each segment is depicting by the statistics of its color distribution. In every segmentation level, every segment is compared with the rest and its similarity image is created, described the probabilistic similarity of each pixel to the segment. As similar to skin detection process, normal probability distribution corresponding to the standard deviation σ_i and mean μ_i of segment S_i are appraising for each channel j = 1, 2, 3 of the Lab color space, and the probability for each image pixel corresponding to this probability is calculated. The product of the probabilities (1) is the final probability is calculated in every channel separately.

$$P_{simIm_{li}}(X) = \prod_{j=1}^{3} \mathcal{N}(X, \mu_{ij}, \sigma_{ij}) \qquad (2)$$

The mask is used for adequate sampling in lieu of torso fitting; therefore, it is calculated as a large square with sides of 2.5PL, with the uppermost side centered in terms of face's center. The mask is applied to every similarity image and its corresponding segment is scored during search process. In proposed system, the approach has the points of interest of considering diverse perceptual groupings and being ready to reduce the requirement for an exact torso estimation, by conjunctively measuring the background and foreground possibilities. Results might be enhanced by including more segmentation levels and masks at various sizes and areas, however the cost of computational multifaceted nature. Here, demonstrate how even with harsh approximations, proposed system accomplish exact and powerful comes about without forcing computational strain.

The undeniable strides are to threshold the amassed potential torso images keeping in mind the end goal to recover the abdominal area mask. Much of the time, hands or arm's skin is not inspected enough amid the torso looking procedure, particularly in the cases, where arms are outstretched. Along these lines, we utilize the skin masks evaluated amid the skin location prepare, which are more exact than for situation they were recovered amid this procedure, since they were computed utilizing the face's skin color, in a color space more suitable for skin and portions made at a better level of segmentation. These are superimposed on the totaled potential torso images and get the most astounding potential.

Rather than utilizing a straightforward or even versatile thresholding, utilize multiple level thresholding to recuperate the districts with solid potential as indicated by the strategy portrayed, however at the same time

conform to the accompanying criteria: 1) they frame a locale estimate near the normal torso size (really greater in request to consider the case, where arms are outstretched), and 2) the external border of this locale covers with adequately high gradients. The distance of the chose locale at threshold t(Region$_t$) to the expected upper body size (ExpUpperBodySize) is computed as takes after:

$$ScoreSize = e^{\dfrac{-|\text{Region}_{t\_}\text{ExpUpperBodySize}|}{\text{ExpUpperbody}}} \qquad (3)$$

Where ExpUpperBodySize = 11 x PL^2. For second criterion the score is calculated by averaging the gradient image (GradIm) responses for the pixels that belong to the perimeter (PRegion$_t$) of Region$_t$ as

$$ScoreGrad = \frac{1}{|PRegion_t|} \sum_{}^{|PRegion_t|} GradIm \cap PRegion_t \quad (4)$$

LOWER BODY SEGMENTATION

The calculation for assessing lower body part, all together to accomplish full body segmentation is fundamentally the same as the one for extraction of upper body. The distinction is the anchor points that start the leg looking procedure. On account of upper body segmentation, it was the position of the face that guides the estimation of the lower body's position. All the more particularly, the general model utilize is that the upper parts of the legs ought to be underneath and close to torso region. In spite of the fact that the already evaluated UBR gives a strong beginning stage to the leg limitation, distinctive sorts of apparel like long coats, dresses, or color likenesses between the garments of the upper and lower body may make the torso area seem distinctive (more often than not longer) than it ought to be. To better gauge the torso region, play out a more refined torso fitting procedure, which does not require broad calculations, since the as of now assessed shape gives a decent guide.

The normal measurements of the torso are again figured in view of anthropometric imperatives, however in a more precise show. Moreover, keeping in mind the end goal to adapt slight body distortions, permit the rectangle to be developed concurring to an obliged parameter space of most noteworthy dimensionality and granularity. In particular, permit revolutions concerning rectangle's center by angle ϕ, translations x- and y- axes, τ_x and τ_y and scaling in x- and y- axes, s_x and s_y. The underlying measurements of the rectangle relate to the normal torso in full frontal and upright view and it is diminished amid looking keeping in mind the end goal to suit different poses. The method of reasoning behind the fitting score of every rectangle is measuring the amount it covers the UBR, since the torso is the biggest semantic locale of the upper body coverage (UBC), while in the meantime covering less of the background region, characterized by potential S (for robustness or solidity). At long last, as a rule, the rectangle should be realigned as for the face's center (FaceCenter) to recoup from misalignments created by various poses and mistakes. A supportive model is the greatest distance of the rectangle's upper corners (LShoulder, RShoulder) from the face's center (D$_{sf}$), which ought to be compelled. In this manner, fitting of the torso rectangle is planned as a boost issue

$$\theta \max f(\theta) = \alpha_1 \times UBC(\theta) + \alpha_2 \times S(\theta) + \alpha_3 \times D_{sf}(\theta) \qquad (5)$$

Where $\theta = (\phi, \tau_x, \tau_y, s_x, s_y)$

$$UBC(\theta) = \frac{\sum TorsoMask(\theta) \cap UBR}{\sum TorsoMask(\theta)}$$

$$S(\theta) = \frac{\sum TorsoMask(\theta)}{\sum UBR}$$

$$Dsf(\theta) = e^{\frac{-|MaxDsf - 1.5 \times PL|}{1.5 \times PL}}$$

Where UBR is the binary image, where pixels inside the UBR are 1, else 0; TorsoMask(θ) is the binary image, where pixels inside the rectangle r$_{TorsoMask}$(θ) are 1, else 0; a_1, a_2, a_3 are weights, set to 0.4, 0.5, and 0.1 individually, and; MasxD_{sf} = max(d(FaceCenter,RShoulder(θ))) and d(FaceCenter, LShoulder(θ)) is the distance of the farthest shoulder point form the face's center.

SPLINE REGRESSION

The problem considers can be depicted as takes after. Given an image I with n pixels to be segmented, namely, I = {p1,…,p$_n$}, and two labeled pixel sets, $F (\subset I)$ and $B(\subset I)$. Here, F contains the client determined foreground

pixels, while B contains the client determined background pixels. The assignment is to dole out a class label in L = {Foreground, Background} to each of the unlabeled pixel $p_i \in I$.

Let $X = \{x_i\}_{i=1}^{n} \subset R^d$ gathered the element vectors of $\{p_i\}_{i=1}^{n}$, where d is the feature dimensionality. Facilitate assume that client labeled n_F foreground pixels and n_B background pixels. Correspondingly, two subsets of features are getting: $u_F = \{x_i^F\}_{i=1}^{n_F} (\subset X)$ and $u_B = \{x_i^B\}_{i=1}^{n_B} (\subset X)$. Presently need to derive the marks of the unlabeled pixels in X.

Before building the spline, assign "+1" to every data points in u_F and "-1" to every data points in u_B as a interpolation values. The end of the goal is to build a spline function F such that for every $x_i^E \in u_F$, $f(x_i^F) \approx 1$ and for every $x_i^B \in u_B$, $f(x_i^B) \approx -1$.

This errand can be considered in a general regularization structure containing function smoothness and data fitting.

$$J(f) = \sum_{i=1}^{n_F}(1 - f(x_i^F))^2 + \sum_{i=1}^{n_B}(-1 - f(x_i^B))^2 + \lambda S(f) \qquad (6)$$

Where S(f) is a smoothness penalty in d dimensions on the function f, and λ is the regularization parameter.
At long last, the regression values of the unlabeled pixels can be straightforwardly got by means of f(x). Since utilize the "+1" as the interpolation value of every foreground pixel and "-1" as that of every background pixel, the median value "0" can be taken as a characterization threshold. Subsequently, for every pixel p_i, its class label can be doled out as takes after:

$$l_i = \begin{cases} \text{Foreground, if } f(x_i) \geq 0 \\ \text{Background, if } f(x_i) \leq 0. \end{cases}$$

SIMULATION RESULTS
To evaluated proposed scheme, utilized samples from the freely accessible INRIA person dataset [], which incorporates persons performing regular exercises in outside situations in for the most part upright position. This is a testing dataset, since the photographs are taken under different brightening conditions, in intensely cluttered environments, persons show up in different sorts of apparel. Proposed scheme implementation based on MATLAB2013a and processing keeps going by and large 3 min for each image with size of 640 x 480 pixels over a machine configuration Intel(R) Pentium(R) CPU B980 @ 2.40GHz, 2GB RAM, 64bit, windows8. The results are shown in following figures:
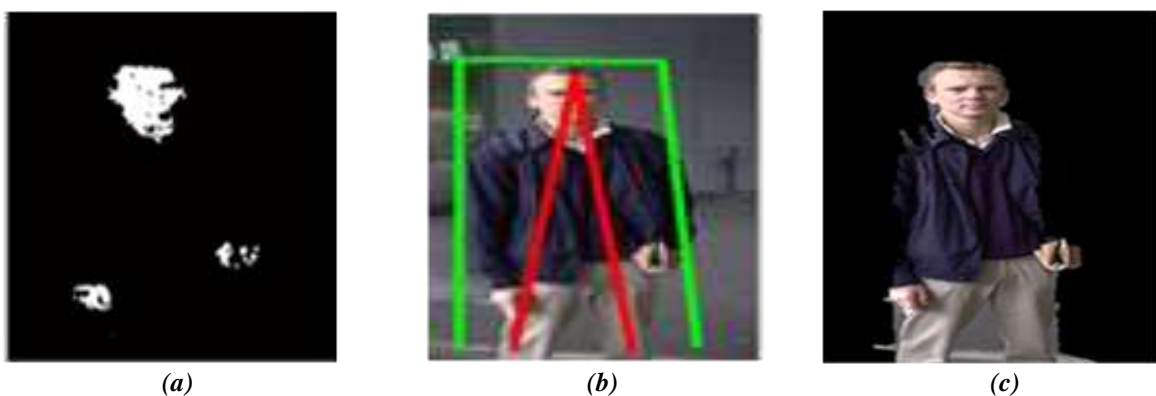

Fig.3. Input Image

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)

(d)　　　　　　　　　　(e)　　　　　　　　　　(f)

Fig.4. Simulation Results for given input. (a) Face and Skin detection, (b) Rectangular mask for upper body detection, (c) Face and Upper body Collaboration, (d) Foreground selection, (e) Background selection, (f) Final Result.

Spline Regression simulation results are shown in following figures:

(a)　　　　　　　　　　*(b)*　　　　　　　　　　*(c)*

Fig.5. Extracting standing human bodies using spline regression. (a) Face and Skin detection, (b) Foreground and background selection using spline detector, (c) Final Result.

CONCLUSION

Proposed system presented a novel procedure for extracting human bodies from single images. It is a bottom up approach that joints data from the multiple levels of segmentation keeping in mind the end goal to find salient regions with high capability of having a place with the human body. The fundamental segment of the framework is the face recognition step, where assess the rough location of the body, develop a rough anthropometric model. Delicate anthropometric requirements direct a productive for the most visible body parts, to be specific the upper

and lower body, maintaining a strategic distance from the requirement for solid earlier learning, for example pose of the body. Probes a testing a dataset appeared that the algorithm can beat best in class segmentation algorithms. Nonetheless, make a few presumptions about the human pose, which limit it from being relevant to uncommon poses also, when impediments are strong. In the future, mean to bargain with more intricate poses, without fundamentally depending on solid pose earlier. Issues like missing extraordinary regions, for example, hair, shoes, and gloves can be tackled by consolidation of more masks in the look for these parts; however alert ought to be taken in keeping the computational unpredictability from rising unnecessarily.

REFERENCES

[1] Athanasios Tsitsoulis, Member, IEEE, and Nikolaos G. Bourbakis, fellow, IEEE, "A methodology for extracting standing human bodies from single images", IEEE transactions on human-machine systems, vol.45, no.3, June 2015.

[2] M. Andriluka, S. Roth, and B. Schiele, "Pictorial structures revisited: People detection and articulated pose estimation," inProc. IEEE Conf. Comput. Vis. Pattern Recog., 2009, pp. 1014–1021.

[3] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge,"Int. J. Comput. Vis., vol. 88, no. 2, pp. 303–338, 2010.

[4] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," inProc. IEEE Conf. Comput. Vis. Pattern Recog., 2008, pp. 1–8.

[5] M. P. Kumar, A. Zisserman, and P. H. Torr, "Efficient discriminative learning of parts-based models," in Proc. IEEE 12th Int. Conf. Comput. Vis., 2009, pp. 552–559.

[6] V. Delaitre, I. Laptev, and J. Sivic, "Recognizing human actions in still images: A study of bag-of-features and part-based representations," in Proc. IEEE Brit. Mach. Vis. Conf., 2010.

[7] A. Gupta, A. Kembhavi, and L. S. Davis, "Observing human-object interactions: Using spatial and functional compatibility for recognition,"IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 10, pp. 1775–1789, Oct. 2009.

[8] B. Yao and L. Fei-Fei, "Grouplet: A structured image representation for recognizing human and object interactions," inProc. IEEE Conf. Comput. Vis. Pattern Recog., 2010, pp. 9–16.

[9] P. Buehler, M. Everingham, D. P. Huttenlocher, and A. Zisserman, "Long term arm and hand tracking for continuous sign language TV broadcasts," inProc. 19th Brit. Mach. Vis. Conf., 2008, pp. 1105–1114.

[10] Shiming Xiang, Peiping Nie, Chunxia Zhang, and Changshui Zhang, Member, IEEE, "Interactive natural image segmentation via spline regression," IEEE transactions on image processing, vol., 18, no. 7, July 2009.