

Tracking and indexing subjects in a crowd

TECHNICAL FIELD

[001] A method for detecting, tracking and indexing subjects in a crowd is disclosed. In particular, subjects of interest to be tracked and indexed are humans in a moving or dynamic crowd such as that in a heavy pedestrian traffic area, including detecting faces and/or heads of the human subjects and tracking their movement as captured in a video feed.

BACKGROUND ART

[002] Video cameras installed for surveillance of heavy pedestrian traffic areas such as in airports, train and bus stations, shopping malls, etc. are only useful if the video feed may be quickly analysed so that objects of interest are identified therefrom and tracked correctly, and finally indexed and stored in a quick retrieval manner from the massive amount of video data received and stored. Otherwise, with only the time and location data of traditional cameras, one needs to know the exact time and location of an event in order to retrieve a video clip for reviewing.

[003] Additional cues such as colour, size and other features of a human

subject of interest may be useful to narrow down the pool of subjects of interest among such a large crowd captured over a length of time. Advances in computerised video analytic technology have made it possible to automatically analyse video content, calculate object features and generate some descriptors to index a video clip to allow for limited searching and retrieval in the future. Some existing systems can even automatically detect objects of interest from the video clip or feed and use the object colour information to index the video clip.

[004] To avoid indexing the same object multiple times, it is necessary to accurately identify and track the objects across multiple frames. Tracking of human subjects of interest in deserted scenarios may not be difficult for areas with few activities. To track individuals in crowded and dynamic scenarios such as in airports, train and bus stations, shopping malls, busy pedestrian crossings, street walkway, etc., presents a bigger challenge and not many technologies are able to do so.

[005] U.S. Published Application No. **US-2009/0041297 (Zhang)** discloses a video analysis module capable of detecting and tracking a human head and estimating a subject's height by considering the region from the head to the ground plane. Its image feature generation method involves scale invariant feature transform (SIFT) for each blob-based human subject to be tracked which matching of the subject requires conditional density

propagation algorithm. Colour information is employed in respect of the subject's ratio of skin tone pixels in the head region for detection purposes. These methods do not adequately deal with tracking in crowd where there are constant occurrences of occlusion of the subject.

[006] U.S. Patent No. **US-6,404,900 (Qian)** provides a system that is more adapted to detecting and tracking human faces in a crowd in its video analysis method using linear Kalman filter. However, its face detection does not cover the head or body. Face detection and tracking alone is inadequate in an unstructured crowd where not everyone faces the camera. In U.S. Published Application No. **US-2003/0048926 (Watanabe)**, an object detected using background subtraction and, using characteristic quantities such as shape, size and texture, determined to be human without provision for occurrences of occlusion. In tracking, this prior art relies on association between consecutive frames of the tracked moving image based on characteristic quantities and behaviour which may not be implemented successfully in a crowded field.

[007] In U.S. Published Application No. **US-2006/0018516 (Massoud)**, background subtraction is also used to detect a moving target. A human target's body segment is first identified into portions and a biometric attribute is then calculated and assigned to each portion and compared against a median colour of corresponding portions of other images for ease

of tracking. An algorithm is used to classify the human subject's activity which basically makes this prior art a system for recognising human activity rather than for capability in identifying and tracking subjects in crowded or heavily occluded situations. The system disclosed in U.S. Patent No. **US-6,658,136 (Brumitt)** attempts to detect and track other moving objects as well as humans using background subtraction. Range information is obtained from a stereo imaging system and therefore requires a continuous sequence of depth maps of the scene.

[008] Tracking using colour information has been attempted in U.S. Patent No. **US-7,068,814 (You)** where a visual tracking of a fast-moving target may be made based on a colour attribute such as average hue in an HIS (hue-saturation-intensity) model of the RGB space. This prior art's 3D colour modelling requires preprocessing to enable it to identify objects by its colour and shape which is not customized to identifying human face and head. A colour attribute-based system is also found in Chinese Patent Publication No. **CN 10-1216885 (Ma)** utilising a skin-colour model and a template-matching method to detect a pedestrian face and using a motion analysis algorithm to estimate the subject's body gravity centre before the face and head could be located. This is impractical for crowded situation where the subject body are heavily occluded. Moreover, it uses linear Kafman filter to track each body rather than colour moments.

SUMMARY OF DISCLOSURE

[009] To overcome the weaknesses in quickly identifying and tracking human subjects in a crowded and fast-moving scenario, the present invention provides a more rigorous method for detecting the human body, and using color moments to identify and track the detected human subject of interest. And a better defined colorimetric attributes can be indexed into database for future forensic analysis.

[010] According to the present invention, there is provided a method for detecting, tracking and indexing at least a human subject in a dynamic crowd comprising a) acquiring image frames of colour video feed from a live video camera or a recorded video source, b) detecting the face and/or head of said human subject using a face and/or head detector, c) estimating a human body region of said human subject based on said detected face and/or head region, d) tracking said human subject by using color moments, e) segmenting said human subject into upper and lower body segments, f) indexing said upper and lower body segments separately based on their colorimetric attributes.

[011] In one aspect of the method, the estimating step (c) is preferably obtained by using calibration information and human body modelling information. The calibration is preferably performed to estimate the

physical height of detected human body region for separate regions of the image frame. The detected human body with unmatched size are discarded. Multiple human subjects may be contemporaneously detected by our method.

[012] In a second aspect, the human subject is indexed once his body segment appears to be least occluded. Preferably, for each frame, a newly detected human subject is matched with a tracking candidate pool. Colour distribution is preferably used for similarity matching. Preferably still, colour moments are used as a biometric attribute to calculate the similarity matching for the human subject being tracked. Once matched, the tracking candidate pool is updated based on matching results. As a specific embodiment, occlusion information from the calibration and colorimetric attribute change consistency is considered during tracking. Preferably, a typical image of the human subject having the least occlusion is saved.

[013] In a third aspect, colour moments are used to build human candidates. The human subject's typical image is preferably used for colour analysis and indexing at the end of tracking. The colorimetric attributes comprise indexes derived from a colorimetric model, including RGB and HSV model, wherein the principal or dominant features of the colours are extracted. Preferably, an un-occluded human subject is further segmented into upper and lower body segments based on colour distribution and body

model, and wherein said upper and lower body segments are indexed separately. The indexed body segments may preferably be stored in a manner for searching and retrieval, and apparatus enabling the implementation of this and the aforesaid may be provided.

BRIEF DESCRIPTION OF THE DRAWINGS

[014] To provide a better understanding of the present invention, the following drawings are attached hereto accompanied by a detailed description thereafter in which:

FIGURE 1 shows a schematic block diagram of the major components of our method;

FIGURE 2 is an illustration of size calibration method used in our system;

FIGURE 3 shows a schematic block diagram of human body segmentation;

FIGURE 4 shows a schematic block diagram of human tracking;

FIGURE 5 shows human body model used in our system;

FIGURE 6 shows a schematic block diagram of color analysis and indexing;
and

FIGURE 7 shows the method in our system to segment upper and lower body.

DETAILED DESCRIPTION OF SPECIFIC EMBODIMENTS

[015] As shown in FIGURE 1, the video feed from a video camera may initially be defined, as an option, in terms of a narrower scope or field of vision than the frame. For example, if a camera is installed to monitor an unstructured crowded environment such as a gantry, main entrance or exit, escalator, etc. a region of interest (ROI) may be predefined such that our detecting, tracking and indexing method only applies to the entrance or exit where everyone must eventually pass through and thus efficiently narrow down the scope of imaging to be analysed at a faster rate and incidences of false alarms may be reduced.

[016] Size calibration of image data is an important process to estimate physical properties of objects and determine the angle in which the camera is viewing the object. A few images with human traverse the space is used in the calibration process. As shown in FIGURE 2, for each scene, at least 6 human references are selected spread the whole scene. For each reference, gravity center (x_i, y_i) , the height of the reference (H_i) in the image are used in calibration process. Calibration is performed by a least squares fitting of a second order polynomial to the data we selected from the references:

$$\min_{a_1 \dots a_6} \sum_i [H_i - (H(x_i, y_i; a_1 \dots a_6))]^2$$

where $a_1 \dots a_6$ are the coefficients of the polynomial. Thus, for each position in

the image, we can predict the relative height of a standing human in the image. Size calibration is used to filter out human subjects with irrational height in the image.

[017] Briefly, our method may be described as comprising 4 main components or modules, i.e.

- (i) a face and/or head detection component wherein the face and/or head of the subject is detected and identified;
- (ii) a body segmentation component wherein the body region of the subject may be estimated and segmented as shown in FIGURE 3;
- (iii) a tracking component wherein the segmented subject may be tracked as shown in FIGURE 4; and
- (iv) a colorimetric attribute analysis and indexing component as shown in FIGURE 5

[018] Intel open source computer vision library provide a robust and real-time face and head detector which can be used in our system. A machine learning approach is used in this object detector. The classifier is trained with face and head samples under different viewing angles and a few critical haar-like features are generated. The different viewing angles of the face and head samples may comprise of the frontal, profile, back view, from -180 degrees to +180 degrees. They may also comprise different tilt angles.

These features are then applied to detect faces and heads in surveillance images. Our system uses both face and head detector to detect the face and head region of a human object of interest since not all humans would be facing the camera in an unstructured crowd. Once a human's face and/or head are detected (please note that this two approach can be independent), based on its size and position in surveillance images, its body region can be estimated using calibration information and human body model. As shown in FIGURE 5, the ratio of head width and body width is 0.47 and the ratio of head height and body height is 0.136 in our system. To further decrease false alarms, size calibration is done to estimate the human body height in surveillance images. Body height with irrational size may be discarded.

[019] Once a human subject is detected based on this face and/or head size and position, and its body is estimated based on the human body model, the size calibration information may be used to detect occlusion between human bodies in a rough manner, i.e. providing basic information as to who is occluding who. If two human subjects are detected in an image frame and these two body regions have a large portion of overlap, the human subject with larger body size and higher face region is tent to occlude the other one. In addition, if a human object is occluded by the other human object, his upper and lower body colorimetric attributes change abruptly. This kind of inconsistently change is also used in our system to detect occlusion. With these occlusion information obtained as input, the human body region may

now be outlined as a segment (a process which we shall call segmentation) for which we may set out to track.

[020] Object tracking is of utmost importance for automatic indexing of video content to eliminate multiple indexing of the same object. In the human tracking component, after the human body has been segmented, the body segment can be used for tracking in a crowded scene. For each human object that is normally tracked, it will be put into the tracking candidate pool. Tracking candidate pool contains all the currently detected human objects. For every newly detected human object, if it is matched with an object inside the candidate pool, the matched object's status will be updated. If the newly detected object is not matched to any object inside the pool, it will be added to the pool. At each frame, newly detected human subjects are matched with a tracking candidate pool. Colour distribution for similarity matching is used. Colour moments are used as a biometric attribute to calculate the matching similarity for human target tracking.

[021] Colour moments compute only the dominant features instead of storing the complete colour distribution. One way to implement this approach is to store the first 2 moments of each colour channel of an image in the index. The 2 moments are (i) average, (ii) variance. For example, for a HSV (hue-saturation-value) image only 6 floating point numbers per

image is stored. The similarity function used for searching, matching and retrieval is a weighted sum of the absolute differences between corresponding moments.

[022] After matching, the tracking candidate pool will be updated based on matching results. Again, the occlusion information based on the calibration is considered during tracking. A typical image with the least occlusion is then saved for each tracking object.

[023] In the fourth and last component, human candidates are built using colour moments which are explained above. As shown in FIGURE 6, when the human object is at the end of track, the body may be further segmented into upper and lower body segments based on color distribution and body model so that the upper and lower body colors may be analyzed and indexed separately and archived into a data storage module for future forensic search and analysis. As shown in FIGURE 7, based on human body region and body model, two segmentation regions can be defined. One is head and torso region and the other is torso and bottom region. For each region, color distribution is analyzed to find the optimized segmentation line. With segmentation lines, human body can be segmented into upper and lower portion. For each portion, both HSV and RGB color models are used to extract the principle colors. Alternatively, other color models may be used. After that, the principle colors for two portions can be indexed. The detailed

colour information thus allow search results to be more accurately matched and narrow down the hit list or length of video clip to be analysed upon retrieval.

[024] While the aforesaid describes a specific embodiment of our method, it would be obvious to a skilled person that many aspects of our invention are capable of being implemented or presented in various alternative ways, or with parts or components substituted, varied or modified accordingly without departing from its working principles, particularly in respect of the use of colour moments for identifying, segmenting and tracking in addition to indexing. For example, apart from RGB and HSV color models, a skilled person would be able to adapt an alternative colorimetric model to be implemented instead, such as the CIE-XYZ model, CMYK model, etc. Such variations are to be considered as falling within the letter and scope of the following claims.

CLAIMS

1. A method for detecting, tracking and indexing at least a human subject in a dynamic crowd comprising
 - a) acquiring image frames of colour¹ video feed from a live video camera or a recorded video source;
 - b) detecting the face and/or head of said human subject using a head/or face detector;
 - c) estimating a human body region of said human subject based on said detected face and/or head region;
 - d) tracking said human subject by using color moments;
 - e) segmenting said human subject into an upper and lower body segments;
 - f) indexing said upper and lower body segments separately based on their colorimetric attributes.
 2. A method according to Claim 1 wherein the human body region estimating step (c) is obtained by using calibration information as well as human body modelling information.
 3. A method according to Claim 2 wherein the calibration is performed to estimate the height of the human body region.
-

4. A method according to Claim 3 wherein the detected human body with irrational human height ranges are discarded.
6. A method according to Claim 1 wherein multiple human subjects are being contemporaneously detected.
7. A method according to Claim 1 wherein tracking step (d) comprises matching a newly detected human subject with a tracking candidate pool for each frame.
8. A method according to Claim 7 wherein the tracking candidate pool comprises all human subjects being tracked.
9. A method according to claim 7 wherein once a newly detected human subject is matched with a tracked human subject in the tracking candidate pool, the status of the tracking candidate pool is updated.
10. A method according to claim 7, wherein if no match is found, the newly detected human subject will be added to the tracking candidate pool.
11. A method according to Claim 7 wherein the tracking candidate pool is updated based on matching results.

12. A method according to any of the previous claims wherein colour moments are used as a biometric attribute to calculate the similarity matching for the human subject being tracked.
13. A method according to any of the previous claims wherein human occlusion is detected based on the calibration information in Claim 3 and colorimetric attributes change consistency.
14. A method according to Claim 13 wherein a typical image of the human subject having the least occlusion is saved.
15. A method according to Claim 14 wherein a typical image of the human subject is used for color analysis and indexing.
16. A method according to Claim 15 wherein step (e) further comprises segmenting said human subject into upper and lower body segments based on color distribution and human body model.

17. A method according to Claim 16 wherein RGB and HSV color models are used in colorimetric attribute derivation for both upper and lower body, separately.

18. A method according to Claim 17 wherein the indexed body segments are stored in a manner for searching and retrieval.

19. An apparatus for implementing a method according to any one of Claims 1 – 18.

20. A system for implementing a method according to any one of claims 1-18.
